STUDIES ON SYSTEM ANALYSIS AND SYNTHESIS
BY
MARKOV RENEWAL PROCESSES

Shunji Osaki

January 1970

STUDIES ON SYSTEM ANALYSIS AND SYNTHESIS

BY

MARKOV RENEWAL PROCESSES

Shunji Osaki

January 1970

STUDIES ON SYSTEM ANALYSIS AND SYNTHESIS
BY
MARKOV RENEWAL PROCESSES

by

Shunji Osaki

Submitted in partial fulfillment of the

requirement for the degree of

DOCTOR OF ENGINEERING

(Applied Mathematics and Physics)

at

KYOTO UNIVERSITY

Kyoto, JAPAN

January 1970

ACKNOWLEDGMENT

The author would like to express his sincere appreciation to Professor Hisashi Mine for his continual encouragement and many helpful suggestions toward the development of the thesis.

The author wishes to express his thanks to Associate Professor Toshiharu Hasegawa for his advice and encouragement. His thanks are also to the members of Professor Mine's Laboratory of Kyoto University, and especially to Mr. Tatsuyuki Asakura and Mr. Toshio Nakagawa for their helpful suggestions and discussions. Finally, the author wishes to thank Miss Noriko Tsuji for typing the manuscript.

ABSTRACT

System analysis and synthesis by Markov renewal processes (which include renewal processes and discrete time Markov chains as special cases) are discussed in this thesis.

Reliability analysis of redundant systems of repairable units are considered and the Lapalace-Stieltjes transforms of the time distributions to first system down are obtained. Signal flow graph method of obtaining the Lapalce-Stieltjes transforms by using Mason's gain formula is also presented. A standby redundant repairable system with preventive maintenance is also presented.

Markovian decision processes and Markov renewal programs are discussed as system synthesis. In Markovian decision processes the policy iteration and linear programming algorithms for the discounted and the nondiscounted models are discussed and the relationship between the two algorithms is also discussed. A new algorithm is proposed from the relationship. Numerical examples and comparison with the three algorithms just mentioned above are presented for Markovian decision processes. The similar discussions are made for Markov renewal programs.

CONTENTS

Chapter III

Chapter IV

Chapter V

Chapter VI

Chapter VII

Chapter VIII

INTRODUCTION

0. 1. Problems of System Analysis and Synthesis

The remarkable progress of engineering techniques yields various kind of systems. The systems are from a simple system of a machine tool to a large-scale system such as the Manned Spacecraft Center in Houston. As a simple example of systems, we consider a system of a machine tool. We shall below describe the problems of system analysis and system synthesis by demonstrating the system of a machine tool.

System Analysis

First we consider the problems of system analysis. For a system of a machine tool, the performance of the system is assumed to be defined as that the machine is operable. If the machine is down, we cannot perform its function. Then we should consider the maintenance problem of the machine. Before the discussion of the maintenance problem, we should know the failure law of the machine. That is, we should investigate the failure time distribution of the machine. The random variables occurring in such problems are all nonnegative. One of the simplest examples is a random failure law, i.e., the exponential failure time distribution. Some of failure time distributions can be further considered: The gamma, the Weibull, the extreme value, the truncated normal, the log normal, and the regular (constant time) distributions. In this thesis, we shall discuss the failure time distributions as the exponential ones or the arbitrary ones. The analysis of systems with the exponential distributions is easy because of the "memoryless property."

To maintain the system, we can consider the following three policies:

(i) The machine is repairable.

(ii) The redundant machines are provided.

(iii) The inspection or the preventive repair is made before failure.

The first policy is that we have repair facilities of the machine. If the machine fails, then the repair of the failed machine is made. The repair time is also random. The repair time distributions can be similarly considered as described in the failure time distributions. After the repair completion the machine recovers its function, i.e., the machine can be operable.

The second policy is the redundancy technique. That is, if two or more machines are provided, we may use the provided machine instead of the failed machine. We can further consider the redundant repairable system, i.e., the system in which the repair of the failed machine is made when any of the other machines are operable. This is a simple redundant repairable model. If the two machines are provided and they are used alternatively, this system is called a two-unit standby redundant model which will be discussed in this thesis. We will further discuss some redundant models.

The third policy is the preventive maintenance one. If the failure time distribution of the machine has Increasing Failure Rate (IFR) [5, p. 12], i.e., the probability of the failure increases as the elapsed time is longer, we should make the inspection or the preventive repair before failure since the inspection or the preventive repair is easier and shorter in time than that of the usual repair.

Systems considered in this thesis are redundant systems of multiple units (or subsystems) each of which is repairable. The "unit" refers to a machine, a computer, a generator, and others. Systems with preventive maintenance are also considered.

In this thesis, we shall consider the situations where the total system failure is a catastrophe. The recent large-scale and complicated systems have such situations. Then our concern of the systems is the time to first system down starting in an initial state. We shall discuss the time to system down throughout Chapters I-V.

System Synthesis

Second we consider the problems of system synthesis. In this thesis we shall discuss Markovian decision processes and Markov renewal programs as the problems of system synthesis. A Markovian decision process is a stochastic sequential decision process based on a discrete time Markov chain. Demonstrating a system of a machine tool just mentioned above, we describe a simple example of Markovian decision processes or Markov renewal programs.

Consider a system of a machine tool. The machine can be operated synchronously, say once an hour. At each period there are two states; one is operating (state 1), and the other is a condition of failure (state 2). If the machine fails, the machine can be restored to perfect functioning by repair. At any period, if the machine is running, we are assumed to earn the return of \$3.00 per period; the probability of being in state 1 at next step is 0.7, the probability of moving to state 2 is 0.3. If the machine is in failure condition, we have two actions to repair the failed machine; one is a rapid repair which requires a cost of \$2.00 (i.e., a return of -\$2.00) with the probability of moving to state 1 of 0.6, another is a usual repair which yields the cost of \$1.00 (i.e., a return of -\$1.00) with the probability of moving to state 1 of 0.4. For the model considered, there are two alternatives available in each state. The state transition diagrams of the alternatives are shown in Fig. 0.1.

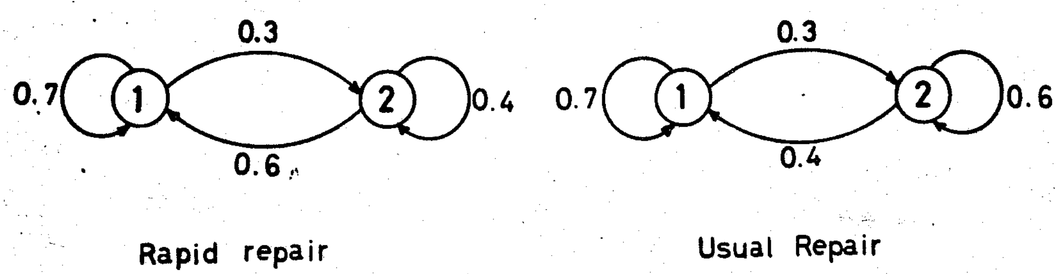

Fig. 0.1 Transition diagrams for the machine maintenance problem.

Under these situations, what strategies we may choose to maximize the total expected return starting in some initial state? We now consider a time planning horizon. If a time planning horizon is finite, then we call the case a "finite horizon problem." A finite horizon problem can be directly formulated by Dynamic Programming and an optimal strategy which is a sequence of decisions for each state can be obtained by means of a recurrence relation. On the other hand, if a time planning horizon is infinite, then we call the case an "infinite horizon problem." In this case, we can not apply either the dynamic programming approach or the direct enumeration approach since the number of the strategies considered is infinite. Furthermore, the total expected return is not necessarily convergent. Then, we consider two approaches for the infinite horizon problem. One is an introduction of a *discount factor* and the other is an introduction of average return per period; a problem of the former type is called a "discounted problem" or a "problem with discounting." If we introduce a discount factor β $(0 \le \beta < 1)$, a unit return at the n period is worth only β^n, and the total expected return is always convergent. For the discounted problem we may consider the total expected return as our objective function. On the other hand, we may consider a "problem with no discounting" or a "nondiscounted problem." For the nondiscounted problem we may consider the average return per period in the steady state as our objective function, if the total expected return is divergent. We can also consider a case where the total expected return is convergent for a nondiscounted problem.

Our main concerns are to find a strategy for maximizing the given objective function, and to determine its maximum value. Such processes are called *Markovian Decision Processes*.

Markovian decision processes are based on discrete time Markov chains. We can extend a discrete time process to a continuous time process. Let us consider an example similar to the preceding one. In the preceding example, we have considered that the machine is operated in discrete time. We can also consider a machine which can be operated continuously in time. If the machine fails, it is repiared by either usual or rapid repair action. The distribution of the running time is $F_{12}(t)$ ($t \geq 0$), while the two distributions of completing a repair are $F_{21}{}^{k}(t)$ ($t \geq 0$), $k = 1$ and 2, respectively, where k denotes any action in state 2. Furthermore, we may consider returns which depend on the duration of the elapsed time in any state. Such a process is based on a Markov renewal process or a semi-Markov process; we call it a *Markov Renewal Program* or a *Semi-Markovian Decision Process*. As a planning horizon we can consider either a finite or an infinite horizon. Here we shall consider only an infinite horizon problem for Markov renewal programs.

0. 2. Review of the Literature

We review the literature on system analysis and synthesis in this section. The bracketed number stands for the reference listed in the end of this thesis.

Many contributions to the reliability theory have been written and a large number of recent papers will be published in technical articles. Barlow and Proschan [5] summarized an excellent book emphasizing the mathematical theory in 1965. In 1965, the Russian mathematicians Gnedenko et al. summarized a book of the reliability theory and it was translated in English [35] in 1969.

The measures of reliability have been defined by many authors and summarized by Barlow and Proschan [5, pp. 5-8]. The measures of reliability are: Reliability, Pointwise availability, Interval availability, Limiting interval availability, Interval reliability, and so on. Hosford [39] has defined three measures of dependability of the system.

The reliability analysis of two-unit redundant systems has been discussed by Epstein and Hosford [26]. Gaver [32, 33] and Liebowitz [49] have discussed a two-unit paralleled (or standby) redundant system. Harris [37] has also discussed a two-unit paralleled redundant system in which the two units are correlated each other. Gnedenko et al. [35] and Srinivasan [58] have discussed a two-unit standby redundant system under the most generalized assumptions. Srinivasan [60] has discussed the same system with noninstantaneous switchover.

Multiple unit redundant systems have been discussed by Barlow [1], Halperin [36], and Srinivasan [59]. Downton [23] has discussed m-out-of-n systems. The reliability analysis for the multiple unit redundant systems by the integral equations of the renewal type has been discussed by Gnedenko [34].

The graphic representation of systems plays an important role in the system theory. In particular, signal flow graphs are applicable to the reliability analysis. The signal flow graphs have been first discussed by Mason [51, 52]. The applications of signal flow graphs are found in Huggins [42, 43]. The relationship between Markov processes and signal flow graphs in the reliability theory has been investigated by Dolazza [22] and Tin Htun [62].

The preventive maintenance theory has been discussed as the replacement problems by Barlow and Hunter [3], Barlow and Proschan [4, 5], and others. Flehinger [12, 13] has discussed some interesting preventive maintenance policies as the marginal checking and marginal testing.

A Markovian decision process was first introduced in 1957 by Bellman [6]. In 1960 Howard [40] published an excellent book in which he discussed some types of Markovian decision processes and gave the policy iteration algorithms for some types of the processes. For the discounted process Blackwell [9] has studied rigorously the process and shown that there exists a β-optimal stationary strategy. Linear programming formulation of the discounted process has been first given by D'Epenoux [21]. De Ghellinck-Eppen [17] have also discussed linear programming formulation. Interesting numerical examples such as the taxicab problem and the automobile replacement problem are found in Howard [40].

Markovian decision processes with no discounting have been also discussed by Howard [40]. Howard [40] has given the policy iteration algorithm for the completely ergodic Markovian decision process and the multichain Markovian decision process. The rigorous discussions for these processes have been studied by Blackwell [9]. Blackwell [9] has also defined a 1-optimal strategy. Veinott [63] has further given the policy iteration algorithm for finding 1-optimal strategies. Optimality of stationary strategies has been given by Derman [20], Blackwell [9], and others. Linear programming formulation has been given by Manne [50], Wolfe-Dantzig [65], and others. For the nondiscounted model the terminating process has been discussed by Howard [40] and Eaton-Zadeh [25].

Markov renewal programs have been first discussed by Jewell [44, 45]. Howard [41] and De Cani [16] have also studied the policy iteration algorithm for the processes. Optimality of stationary strategies has been given by Denardo [18] for the discounted model and by Fox [31] for the nondiscounted model. Denardo-Fox [19] have further studied the multichain Markov renewal programs.

Though we omit in this thesis, the following models are of interest: Markovian decision processes with state and action spaces having nonempty Borel sets have been discussed by Blackwell [10], Strauch [61], and others. Brown [11] has considered Markovian decision processes from the viewpoint of dynamic programming [7, 8]. Contraction mappings in sequential decision processes have been discussed by Denardo [18]. Stochastic games are also of interest (see Shapley [55] and Hoffman-Karp [38]). Applications of Markovian decision processes are found in Klein [48] and others.

In the final part of this section we review mathematical tools used in this thesis. Renewal processes are of importance throughout this thesis. The theory of renewal processes are summarized in Smith [57], Cox [14], and Feller [28]. Discrete time Markov chains are summarized in Kemeny-Snell [47] and Feller [27]. Markov renewal processes, which are extensions of renewal processes and Markov processes, play an important role throughout this thesis. For Markov renewal processes, Smith [56], Pyke [53, 54], and Barlow [2] have discussed in detail. Linear programming theory is found in Dantzig [15].

0. 3. Outline of the Thesis

This thesis is devided into Introduction, Chapters I-VIII, Conclusion, and References. Here we describe the outline of each chapter.

Chapter I discusses the reliability analysis for two-unit redundant systems by integral equations of renewal theory. The three types of two-unit standby (or paralleled) redundant systems will be investigated. Applying the integral equations of renewal theory, we shall analyse each of the three systems.

Chapter II discusses the reliability analysis for redundant systems by integral equations of the renewal type. This chapter studies reliability analysis for multiple unit redundant systems of dissimilar units by using integral equations of the renewal type. Some special models are also discussed.

Chapter III discusses the signal flow graph analysis for systems. The relationship between Markov renewal processes and signal flow graphs is investigated and some examples of the signal flow graph analysis for redundant systems are presented.

Chapter IV discusses a two-unit standby redundant system with standby failure. Taking account of the failure of a standby unit, we shall derive the Laplace-Stieltjes transform of the time distribution to first system down and its mean time. The analysis is made by using the signal flow graph method obtained in the preceding chapter.

Chapter V discusses a two-unit standby redundant system with repair and preventive maintenance. Considering the repair and preventive maintenance policies for a two-unit standby redundant system, we shall obtain the Laplace-Stieltjes transform of the time distribution to first system down and its mean time. The analysis is also made by using the signal flow graph method obtained in Chapter III.

Chapter VI discusses Markovian decision processes with discounting. Definitions and notations of Markovian decision processes are introduced. The policy iteration algorithm is discussed and the linear programming formulation is studied. The relationship between the above two algorithms is discussed and numerical examples are presented.

Chapter VII discusses Markovian decision processes with no discounting. Some properties of the processes are provided and the policy iteration algorithm is discussed for the completely ergodic process. Linear programming formulation is also presented for the completely ergodic process. The relationship between the above two algorithms are discussed and a new algorithm is proposed. Comparison among the above three algorithms is demonstrated for the automobile replacement problem. The terminating process is also discussed. Linear programming considerations for the multichain process is finally presented.

Chapter VIII discusses Markov renewal programs. Markov renewal processes are defined and some useful properties are provided for the preliminaries. Markov renewal processes with returns are also discussed for the discounted case, the nondiscounted case, and the terminating case. Markov renewal programs are introduced. For the nondiscounted model the policy iteration and the linear programming algorithms are discussed and the relationship between the above two algorithms is investigated. We shall also discuss the two algorithms for the nondiscounted model and the terminating model.

Conclusion summarizes the results of this thesis and describes further problems of system analysis and synthesis. Publications list of the author are presented.

References are provided in the end of this thesis.

CHAPTER I

RELIABILITY ANALYSIS FOR TWO-UNIT REDUNDANT SYSTEMS BY INTEGRAL EQUATIONS OF RENEWAL THEORY

1. 1. Introduction

Reliability analysis for a redundant repairable system is important and basic in the theory of reliability. Many contributions to the reliability analysis for such a system have been made by Gaver [32, 33], Downton [23], Gnedenko [34], Srinivasan [58, 59, 60], and others. In this chapter, we shall only discuss two-unit redundant repairable models. A two-unit redundant repairable model is basic and applicable to many practical fields. We shall analyse the following three models:

(i) A two-unit paralleled redundant system.

(ii) A two-unit standby redundant system.

(iii) A two-unit standby redundant system with priority.

Here we assume that the switchover time is instantaneous for models (i), (ii), and (iii). In this chapter, we shall analyse these models systematically and elegantly by using the integral equation of renewal theory [14].

Consider a system composed of two units as a general model described above. Appropriately labeling the number of units, we may call unit 1 and unit 2. We assume that unit i (i = 1, 2) obeys its failure time distribution $F_i(t)$ ($t \geq 0$) and its repair time distribution $G_i(t)$ $(t \geq 0)$. These distributions are mutually independent. We further assume that the repair of a unit recovers its function perfectly. The switch-over time from the failure to the repair, from the repair completion to the operating state (or standby state), or from the standby state to the operating state is supposed to be instantaneous. Necessary assumptions shall be imposed for each model therein. Our concern is the first time to a state that two units are under failure or repair simultaneously starting the initial state that two units are operating (or standby) at $t = 0$. We shall show in this chapter that the analysis for the above models can be made systematically and elegantly by using the integral equation of renewal theory, and we obtain easily the Laplace-Stieltjes (LS) transform of the first time distribution to system down and its mean time for each model.

The renewal theoretic method described in this chapter may be easier than the earlier conventional methods. This method is applicable to other modified systems.

1. 2. Integral Equation of Renewal Theory

This preliminary section describes the integral equation of renewal theory by introducing the concept "cycle." A cycle is the duration of time and the cycle is completed by the realization of an event. An event is defined for each model. For example, a two-unit paralleled redundant system has the cycle which is defined by the time duration from two units operating - one unit failure - repair to repair completion and it is free from system down. Fig. 1.1 shows an example of the cycle for model (i). As soon as the first cycle

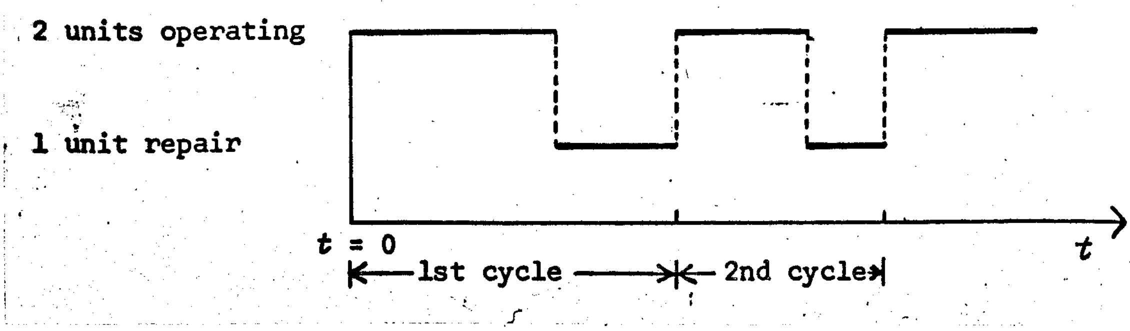

Fig. 1. 1. The behavior of model (i).

is completed, the second cycle begins and so on. For each cycle, the probability of the occurrence of system down exists. Thus we have two possibilities for each cycle: One is the completion of the cycle, and another is the occurrence of system down. These two events are mutually excluseve. The former proceeds in the next cycle, and the latter stops the system behavior.

Let $B(t)$ $(t \geq 0)$ define the probability that {the time duration of each cycle is $\leq t$ and it is free from system down}. Let $A(t)$ $(t \geq 0)$ define the probability that {the time duration to system down for each cycle is $\leq t$}. When each cycle is completed, we assume that the system is the same state at the initial time $t=0$. Let $\Phi(t)$ $(t \geq 0)$ denote the distribution of the first time to system down up to time t. Then we can consider two cases for $\Phi(t)$: One is the occurrence of system down for the first cycle, and another is the completion of the first cycle, which proceeds in the second cycle. Assuming that the second cycle also obeys $\Phi(t)$, we have

$$(1.1) \qquad \Phi(t) = A(t) + B(t) * \Phi(t),$$

where $*$ denoted the convolution operation. Equation (1.1) can also be written by

$$(1.2) \qquad \Phi(t) = \sum_{n=0}^{\infty} [B(t)^{n*}] * A(t),$$

where

$$(1.3) \qquad B(t)^{n*} = \begin{cases} \overbrace{B(t) * B(t) * \cdots * B(t)}^{n} & (n \geq 1) \\ H(t) \text{ (Heaviside step function; } n=0). \end{cases}$$

Equation (1.2) can be interpreted that the occurrence of system down is the first cycle ($A(t)$), the second cycle ($B(t) * A(t)$), the third cycle ($B(t) * B(t) * A(t)$), or so on. and these events are mutually exclusive.

Let the LS transforms† of $A(t)$, $B(t)$, and $\Phi(t)$ denote the corresponding small letters. For example, we define

$$(1.4) \qquad \varphi(s) = \int_0^{\infty} e^{-st} \, d\Phi(t).$$

Applying the LS transforms for (1.1) (or (1.2)), we have

---

† The Laplace-Stieltjes (LS) transform is not the same as the Laplace (L) transform. Where both exist, $LS\{\Phi(t)\} = L\{d\Phi/dt\}$.

(1.5) $\varphi(s) = \alpha(s)/[1 - \beta(s)].$

To verify that $\Phi(t)$ is a proper distribution [28, p. 129] (i.e., $\Phi(\infty) = 1$), we have from (1.5) the following condition:

(1.6) $\varphi(0) = \alpha(0)/[1 - \beta(0)] = 1,$

or

(1.7) $\alpha(0) + \beta(0) = 1.$

Applying the results just mentioned above, we shall obtain the LS transform of the first time distribution to system down. That is, definding the cycle and finding $\alpha(s)$ and $\beta(s)$ for each model, we can obtain $\varphi(s)$ from (1.5). Then we should verify that (1.7) holds so that $\Phi(t)$ is a proper distribution.

1. 3. A Two-Unit Paralleled Redundant System

This section describes a two-unit paralleled redundant system. Since this model depends on the history of the failure time, we assume that the failure time distribution has "memoryless" property [27, p. 411]. That is, we assume

(1.8) $F_i(t) = 1 - \exp(-\lambda_i t). \quad (i = 1, 2)$

we can consider an arbitrary distribution $G_i(t)$ (i = 1, 2). The behavior of the model can be interpreted in detail by Gaver [32, 33] and Liebowitz [49].

As we have mentioned in the preceding section, the model has the cycle which is defined by the time duration from two units operating - one unit failure - repair completion and it is free from system down (see Fig. 1.1), where we assume that two units are operating simultaneously at $t = 0$.

Let's determine $\alpha(s)$. For $\alpha(s)$, by using that unit i ($i = 1, 2$) fails with probability $\lambda_i/(\lambda_1 + \lambda_2)$ and its distribution $1 - e^{-(\lambda_1+\lambda_2)t}$ (i.e., the LS transform $(\lambda_1 + \lambda_2)/(s + \lambda_1 + \lambda_2)$) and then the remaining unit j ($j \neq i$) fails before the repair completion of the failed unit i, we have

$$(1.9) \qquad \alpha(s) = \sum_{i=1}^{2} \frac{\lambda_i}{\lambda_1+\lambda_2} \cdot \frac{\lambda_1+\lambda_2}{s+\lambda_1+\lambda_2} \int_0^\infty e^{-st} d_t \left[\int_0^t \overline{G}_i(t) \lambda_j e^{-\lambda_j t} dt \right]$$

$$= \sum_{i=1}^{2} \frac{\lambda_i}{s+\lambda_1+\lambda_2} \cdot \frac{\lambda_j}{s+\lambda_j} \left[1 - g_i(s+\lambda_j) \right], \quad (j \neq i)$$

where

$$(1.10) \qquad \overline{G}_i(t) = 1 - G_i(t). \qquad (i = 1, 2)$$

For $\beta(s)$, by using that the repair of the failed unit i is completed before the remaining unit j ($j \neq i$) fails, we have

$$(1.11) \qquad \beta(s) = \sum_{i=1}^{2} \frac{\lambda_i}{\lambda_1+\lambda_2} \cdot \frac{\lambda_1+\lambda_2}{s+\lambda_1+\lambda_2} \int_0^\infty e^{-st} d_t \left[\int_0^t e^{-\lambda_j t} dG_i(t) \right]$$

$$= \sum_{i=1}^{2} \frac{\lambda_i}{s+\lambda_1+\lambda_2} g_i(s+\lambda_j). \qquad (j \neq i)$$

Noting that the failure time distributions are memoryless, we have the same state at time $t = 0$ when the first cycle is completed.

From (1.5), we have the LS transform of the first time distribution to system down:

$$(1.12)\quad \varphi(s) = \frac{\sum_{i=1}^{2} \frac{\lambda_i \lambda_j}{s+\lambda_i}[1 - g_i(s+\lambda_j)]}{s+\lambda_1+\lambda_2 - \sum_{i=1}^{2} \lambda_i g_i(s+\lambda_j)} . \qquad (j \neq i)$$

To verify that $\Phi(t)$ is a proper distribution, we have

$$(1.13)\quad \alpha(0) + \beta(0) = \sum_{i=1}^{2} \frac{\lambda_i}{\lambda_1+\lambda_2} g_i(\lambda_j) + \sum_{i=1}^{2} \frac{\lambda_i}{\lambda_1+\lambda_2}[1 - g_i(\lambda_j)] = 1.$$

For the mean time to system down, we have

$$(1.14)\quad \hat{T} = -\left.\frac{d\varphi(s)}{ds}\right|_{s=0} = \frac{1 + \sum_{i=1}^{2} \frac{\lambda_i}{\lambda_j}[1 - g_i(\lambda_j)]}{\sum_{i=1}^{2} \lambda_i [1 - g_i(\lambda_j)]} . \quad (j \neq i)$$

These results (1.12) and (1.14) have been given by Gaver [33].

Gaver [33] has further considered the overloading situation. That is, when unit i ($i = 1, 2$) fails, the remaining unit j ($j \neq i$) may be overloaded. We have two situations on overloading. One is the probability α_j that the system immediately fails because the remaining unit j is overloaded when unit i is under repair. Another is the probability $\bar{S}_j(t) = 1 - S_j(t)$ that the remaining unit j can perform its function without system down for the time duration at least t starting from the instant that the remaining unit j is operating. This modified model can be easily obtained in the same fashion. For $\alpha(s)$ and $\beta(s)$, we have

$$(1.15)\quad \alpha(s) = \sum_{i=1}^{2} \frac{\lambda_i}{s+\lambda_1+\lambda_2}\Big[\alpha_j + (1-\alpha_j)\int_0^\infty e^{-st}\lambda_j e^{-\lambda_j t}\bar{G}_i(t)\bar{S}_j(t)\,dt$$

$$+ (1-\alpha_j)\int_0^\infty e^{-st}e^{-\lambda_j t}\bar{G}_i(t)\,dS_j(t)\Big], \quad (j \neq i)$$

and

$$(1.16)\quad \beta(s) = \sum_{i=1}^{2} \frac{\lambda_i}{s+\lambda_1+\lambda_2}(1-\alpha_j)\int_0^\infty e^{-st}e^{-\lambda_j t}\bar{S}_j(t)\,dG_i(t). \quad (j \neq i)$$

Substituting (1. 15) and (1.16) into (1.5), we can obtain $\varphi(s)$. We further can obtain its mean time. Here we omit these results (see equations (24) and (10) in Gaver [33]).

1. 4. A Two-Unit Standby Redundant System

This section describes a two-unit standby redundant system, where the switchover time is instantaneous. The model can be analysed under the most generalized assumption that the failure and repair time distributions are arbitrary.

Initially at $t = 0$, unit 1 is operating and unit 2 is in standby. As soon as unit 1 fails, unit 2 is operating and unit 1 undergoes repair. When the repair of unit 1 is completed befor unit 2 fails, unit 1 is in standby and then as soon as unit 2 fails, unit 1 in standby begins to be operating. While we consider another case that unit 2 fails before the repair completion of unit 1, which implies system down. The system behaves from the operating unit 2 to the operating unit 1 and so on if it is free from system down.

To use the "cycle" mentioned above, we should change the initial time. That is, the initial time is the instant that unit 2 begins to be operating and unit 1 undergoes repair in the first time. The time distribution from the old initial time to the new has $F_1(t)$. The desired LS transform $\varphi(s)$ can be given by

$$\varphi(s) = f_1(s)\,\varphi_1(s), \tag{1.17}$$

where $f_1(s)$ is the LS transform of $F_1(t)$ and $\varphi_1(s)$ is the LS transform under the new initial time.

For $\varphi_1(s)$, we define the cycle as follows: The cycle is the time duration from the instant that unit 2 begins to be operating (unit 1 begins to repair) to the instant that unit 2 begins again to be operating via failure-repair-standby of unit 2 and failure-repair-standby of unit 1 (see Fig. 1.2). The occurrence of

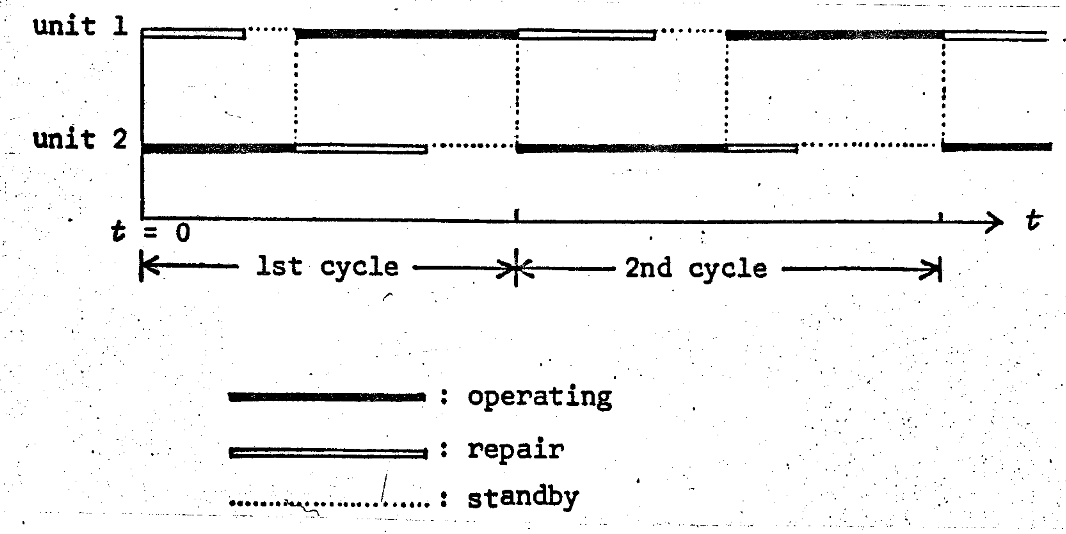

Fig. 1. 2. The behavior of model (ii) (for $\varphi_1(s)$).

system down has two cases for each cycle: The first is that unit 2 fails before the repair completion of unit 1, whose distribution is $\int_0^t \overline{G}_1(t)\,dF_2(t)$ in the same way of (1.9). The second is that the repair of unit 1 is completed (and thereon unit 1 undergoes standby state) before unit 2 fails, after the failure of unit 2, unit 1 in standby begins to be operating and unit 1 fails before the repair completion of unit 2.

Let's consider the probability that the repair of unit 1 is completed (and thereon unit 1 undergoes standby state) up to time t with $G_1(t)$ and unit 2 fails in the interval $(t,\ t + dt)$ with $dF_2(t)$. Thus we have $\int_0^t G_1(t)\,dF_2(t)$.

These two cases are mutually exclusive and the second case has the convolution of the distributions (i.e., the product of the LS transforms) of $\int_0^t G_1(t)\,dF_2(t)$ and $\int_0^t \overline{G}_2(t)\,dF_1(t)$. For $\alpha(s)$, we have

$$(1.18)\quad \begin{aligned}\alpha(s) &= \int_0^\infty e^{-st} d_t\left[\int_0^t \overline{G}_1(t)dF_2(t)\right] \\ &\quad + \int_0^\infty e^{-st} d_t\left[\int_0^t G_1(t)dF_2(t)\right]\int_0^\infty e^{-st} d_t\left[\int_0^t \overline{G}_2(t)dF_1(t)\right] \\ &= \int_0^\infty e^{-st}\overline{G}_1(t)dF_2(t) + \int_0^\infty e^{-st}G_1(t)dF_2(t)\int_0^\infty e^{-st}\overline{G}_2(t)\,dF_1(t).\end{aligned}$$

For $\beta(s)$, the cycle is the time duration from the instant that unit 2 begins to be operating to the instant that unit 2 begins again to be operating. Thus we have

$$(1.19)\quad \beta(s) = \int_0^\infty e^{-st}G_1(t)\,dF_2(t)\int_0^\infty e^{-st}\,G_2(t)\,dF_1(t).$$

Using (1.18) and (1.19), we have

(1.20) $$\varphi_1(s) = \alpha(s)/[1-\beta(s)].$$

From (1.17), we have

(1.21) $$\varphi(s) = f_1(s)\,\varphi_1(s)$$

$$= \frac{f_1(s)\int_0^\infty e^{-st}\overline{G}_1(t)\,dF_2(t) + f_1(s)\int_0^\infty e^{-st}G_1(t)\,dF_2(t)\int_0^\infty e^{-st}\overline{G}_2(t)\,dF_1(t)}{1 - \int_0^\infty e^{-st}G_1(t)\,dF_2(t)\int_0^\infty e^{-st}G_2(t)\,dF_1(t)}.$$

We can easily verify that (1.17) holds for the model.

The mean time is given by

(1.22) $$\hat{T} = -\left.\frac{d\varphi(s)}{ds}\right|_{s=0}$$

$$= \frac{1}{\lambda_1} + \frac{\dfrac{1}{\lambda_2} + \dfrac{1}{\lambda_1}\int_0^\infty G_1(t)\,dF_2(t)}{1 - \int_0^\infty G_1(t)\,dF_2(t)\int_0^\infty G_2(t)\,dF_1(t)},$$

where

(1.23) $$1/\lambda_i = \int_0^\infty t\,dF_i(t). \qquad (i = 1,\ 2)$$

These results (1.21) and (1.22) have been found in Srinivasan [58].

As a special case, we consider that

(1.24) $$F_i(t) = 1 - \exp(-\lambda_i t). \qquad (i = 1,\ 2)$$

In this case (1.21) and (1.22) become

(1.25) $$\varphi(s) = \frac{\dfrac{\lambda_1}{s+\lambda_1}\cdot\dfrac{\lambda_2}{s+\lambda_2}[1-g_1(s+\lambda_2)] + \dfrac{\lambda_1}{s+\lambda_1}\cdot\dfrac{\lambda_2}{s+\lambda_2}\,g_1(s+\lambda_2)\dfrac{\lambda_1}{s+\lambda_1}[1-g_2(s+\lambda_1)]}{1 - \dfrac{\lambda_2}{s+\lambda_2}\,g_1(s+\lambda_2)\dfrac{\lambda_1}{s+\lambda_1}\,g_2(s+\lambda_1)},$$

and

$$(1.26) \qquad \hat{T} = \frac{1}{\lambda_1} + \frac{\frac{1}{\lambda_2} + \frac{1}{\lambda_1} g_1(\lambda_2)}{1 - g_1(\lambda_2) g_2(\lambda_1)} .$$

Gaver [32] has given these results (1.25) and (1.26), where two units are identical, i.e., $\lambda_1 = \lambda_2 = \lambda$ and $G_1(t) = G_2(t) = G(t)$ ($g_1(s) = g_2(s) = g(s)$).

1. 5. A Two-Unit Standby Redundant System with Priority

This section describes a two-unit standby redundant system with priority. This system is composed of two units. Initially at $t = 0$, unit 1 begins to be operating and unit 2 is in standby. When unit 1 fails, unit 2 in standby begins to be operating and unit 1 undergoes repair immediately. When the repair of unit 1 is completed before unit 2 fails, unit 1 begins to be operating again and unit 2 undergoes standby state immediately. While, when unit 2 fails before the completion of the repair of unit 1, system down occurs. For the model, we want to use unit 1 if it is available. So we give the priority for unit 1. Then we have no repair facility for unit 2 because unit 2 stops its operating state on the way and it is free from failure.

For the model, we assume that unit 1 has priority. Unit 1 has its failure time distribution $F_1(t)$ (arbitrary) and its repair time distribution $G_1(t)$ (arbitrary). Unit 2 has its failure time distribution $F_2(t) = 1 - \exp(-\lambda_2 t)$ and no repair facility. We assume that the failure time of unit 2 has "memoryless" property for the convenience of analysis because unit 2 stops its operating state.

For the model, we define the cycle as follows: The cycle is the time duration from the instant that unit 1 begins to be operating (unit 2 is in standby) to the instant that unit 1 begins again to be operating via failure-repair of unit 1.

For $\alpha(s)$, we can consider that when unit 1 fails, unit 2 begins to be operating and unit 2 fails before the repair completion of unit 1. The failure time of unit 1 obeys $F_1(t)$ (its LS transform $f_1(s)$) and the distribution that unit 2 fails before the repair completion of unit 1 at least t obeys $\int_0^t \overline{G_1}(t)\lambda_2 e^{-\lambda_2 t}dt$. Thus we have

$$(1.27)\qquad \alpha(s) = f_1(s)\int_0^\infty e^{-st} d_t\left[\int_0^t \overline{G_1}(t)\lambda_2 e^{-\lambda_2 t}dt\right]$$

$$= f_1(s)\frac{\lambda_2}{s+\lambda_2}\left[1-g_1(s+\lambda_2)\right].$$

For $\beta(s)$, we can consider that when unit 1 fails, unit 2 begins to be operating and the repair of unit 1 is completed before unit 2 fails. The distribution that the repair of unit 1 is completed before unit 2 fails at least t obeys $\int_0^t e^{-\lambda_2 t} dG_1(t)$. Thus we have

$$(1.28)\qquad \beta(s) = f_1(s)\int_0^\infty e^{-st} d_t\left[\int_0^t e^{-\lambda_2 t} dG_1(t)\right]$$

$$= f_1(s)\, g_1(s+\lambda_2).$$

From (1.5), we have

$$(1.29) \qquad \varphi(s) = \frac{f_1(s)\frac{\lambda_2}{s+\lambda_2}[1 - g_1(s+\lambda_2)]}{1 - f_1(s)g_1(s+\lambda_2)}.$$

It is evident that (1.7) holds for the model.

The mean time is given by

$$(1.30) \qquad \hat{T} = -\frac{d\varphi(s)}{ds}\bigg|_{s=0} = \frac{1}{\lambda_2} + \frac{1}{\lambda_1[1 - g_1(\lambda_2)]},$$

where

$$(1.31) \qquad 1/\lambda_1 = \int_0^\infty t\,dF_1(t) = -\frac{df_1(s)}{ds}\bigg|_{s=0}.$$

The mean time (1.30) coincides with (1.26) of a two-unit standby redundant system with no priority, where we assume that $\lambda_1 = \lambda_2 = \lambda$ and $G_1(t) = G_2(t) = G(t)$ ($g_1(s) = g_2(s) = g(s)$). This fact is interesting. But we can understand that the mean time $\hat{T}$ in (1.26) has exponential failure time distributions and thus "memoryless" property is independent of "priority".

1. 6. Conclusion

Reliability analysis for two-unit redundant systems has been discussed by using the integral equation of renewal theory and its associated "cycle". The LS transform of the first time distribution to system down for each model and its mean time have been given by defining the cycle, finding $\alpha(s)$ and $\beta(s)$, and using (1.5). The method mentioned in this chapter is elegant and applicable to other modified systems.

The conventional analysis for such systems will be used. As an example we consider the supplementary variable method (see e.g., Gaver [32] and Liebowitz[49]). The supplementary variable method for model (i) is tedious and complicated for calculation. The method described in this chapter seems to be elegant and easy.

Renewal theoretic argument will occur in the field of reliability and in the other areas of systems analysis. The basic concept of integral equation of renewal theory is applicable to many systems and the method mentioned in this chapter will be capable of analysing such systems.

CHAPTER II

RELIABILITY ANALYSIS FOR SYSTEMS BY INTEGRAL EQUATIONS OF THE RENEWAL TYPE

2. 1. Introduction

In this chapter, we consider a system which is composed of several repairable units. We find the mean time to system failure and the failure time distribution for the system, according to a particular criterion of system failure.

We treat the multiple unit system, where each unit is dissimilar. This has two advantages. 1) In practical situations, the units are not identical. 2) Our results for dissimilar units imply many earlier ones as special cases (as shown in Section 2.4). Thus, the results are practical and useful.

Many studies have been made for multiple-unit systems with repair. Epstein-Hosford [26], Barlow [1], and Halperin [36] have discussed such systems with exponential failure and exponential repair. Further, Gaver [32, 33], Downton [23], and Liebowitz [49] have discussed such systems with exponential failure and general repair. In this chapter, we discuss a system with exponential failure and general repair.

In Section 2.2, we consider a 2-out-of-n system in which the failure time distribution of each unit is exponential and the repair time distribution of each unit is general. A 2-out-of-n system is defined to be down if and only if two or more units are simultaneously in a failed state. Then, the Mean Time to System Failure (MTSF) is derived. Further, we derive the MTSF for the system under the overloading condition. In Section 2.3, we show that the failure time distribution for the system is derived from an integral equation of the renewal type. The failure time distribution for the system under the overloading condition is derived in a similar way. In Section 2.4, we show that our results imply many earlier results, e.g., Gaver, Downton and others, as special cases.

2. 2. Mean Failure Times

In this section, we derive the MTSF for a 2-out-of-n system of dissimilar units without first getting the failure-time distribution. It is also derived in the next section from the failure-time distribution.

Consider a system of n units numbered consecutively: $i = 1, 2, \ldots, n$. (The index i will always go from 1 to n; where used as an index for summation, the sum is from 1 to n.) Assume that

(1) the failure time of the ith unit is distributed exponentially with parameter λ_i:

$$F_i(t) \equiv 1 - e^{-\lambda_i t}, \quad t \geq 0, \tag{2.1}$$

(2) failures of the units are mutually independent,

(3) the repair time of the ith unit obeys an arbitrary distribution $G_i(t)$, where the

operation of each unit is fully restored upon repair

We now define

$$(2.2) \qquad \Lambda \equiv \sum_i \lambda_i ,$$

and

$$(2.3) \qquad \lambda_i^* \equiv \Lambda - \lambda_i ;$$

where Λ denotes the parameter of the exponential distribution when one of n operable units fails, and λ_i^* denotes the parameter of the exponential distribution when a unit i is under repair and one of the remaining (n - 1) operable units fails.

We define the states of the system as follows:

state 0: n units are operable,

state i: a unit i is under repair and the remaining (n - 1) units are operable,

state n + 1: two units are down (system failure).

The MTSF for a 2-out-of-n system when the system starts with all n operable units is equivalent to the mean first passage time from state 0 to state n + 1. Let T_j (j = 0, 1, 2, ..., n) denote the mean first passage time from state i to state n + 1. Since the system fails with the parameter Λ and the conditional probability that a unit i fails is λ_i/Λ, we have

$$(2.4) \qquad T_0 = 1/\Lambda + \sum_i (\lambda_i/\Lambda) T_i .$$

Then there are two possible transitions from state i. These are: (1) transition to state 0, i.e., a repair is completed before the remaining (n - 1) units fail in the interval t, and (2) transition to state n + 1, i.e., at least one of the remaining (n - 1) units fails before a repair is completed in that interval. These two events are mutually exclusive. Thus we have

$$(2.5) \qquad T_i = \int_0^\infty (t+T_0)\, e^{-\lambda_i^* t}\, dG_i(t) + \int_0^\infty t\, \overline{G}_i(t)\, \lambda_i^* e^{-\lambda_i^* t}\, dt,$$

where $\overline{G}_i(t) \equiv 1 - G_i(t)$, the complementary distribution. Integrating by parts and rearranging (2.5), we have

$$(2.6) \qquad T_i = [1 - g_i(\lambda_i^*)]/\lambda_i^* + T_0\, g_i(\lambda_i^*),$$

where

$$(2.7) \qquad g_i(\lambda_i^*) \equiv \int_0^\infty e^{-\lambda_i^* t}\, dG_i(t),$$

i.e., the Laplace-Stieltjes (LS) transform of $G_i(t)$ evaluated at $s = \lambda_i^*$. Solving (2.4) and (2.6) for T_0, we have

$$(2.8) \qquad T_0 = \left\{1 + \sum_i \frac{\lambda_i}{\lambda_i^*}[1 - g_i(\lambda_i^*)]\right\} \Big/ \sum_i \lambda_i [1 - g_i(\lambda_i^*)],$$

which is the MTSF for a 2-out-of-n system when the system starts with all n operable units.

In the above model, though one of n units is under repair, the system is assumed to be able to perform its function perfectly. In practical situations, if one unit is under repair, the system may fail because the remaining (n - 1) units are overloaded. So let $A_i(t)$ denote the probability that the system fails up to time t because of overloading when a unit i is under repair. Then, $\bar{A}_i(t) \equiv 1 - A_i(t)$ denotes the probability that the system can perform its function at time t.

For the system under the overloading condition, equation (2.4) also holds. But we have to consider the overloading condition for state i. There are three possible transitions from state i: (1) a transition to state 0 without overloading, (2) a transition to state n + 1 because of overloading, and (3) a transition to state n + 1 because of a failure of the remaining n - 1 units. These three events are mutually exclusive. Thus, we have

$$(2.9) \quad T_i = \int_0^\infty (t + T_0) e^{-\lambda_i^* t} \bar{A}_i(t)\, dG_i(t) + \int_0^\infty t e^{-\lambda_i^* t} \bar{G}_i(t)\, dA_i(t) + \int_0^\infty t\, G_i(t) A_i(t) \lambda_i^* e^{-\lambda_i^* t} dt.$$

Integrating by parts and rearranging (2.9), we have

$$(2.10) \quad T_i = T_0 \int_0^\infty \bar{A}_i(t) e^{-\lambda_i^* t} dG_i(t) + \int_0^\infty \bar{A}_i(t) \bar{G}_i(t) e^{-\lambda_i^* t} dt.$$

Solving (2.4) and (2.10) for T_0, we have

$$(2.11) \quad T_0 = \frac{1 + \sum_i \lambda_i \int_0^\infty \bar{A}_i(t) \bar{G}_i(t) e^{-\lambda_i^* t} dt}{\sum_i \lambda_i \{1 - \int_0^\infty \bar{A}_i(t) e^{-\lambda_i^* t} dG_i(t)\}}.$$

T_0 is the MTSF for a 2-out-of-n system under the overloading condition when the system starts with all n operable units.

2. 3. Failure Time Distributions

In the preceding section, the MTSF for a 2-out-of-n system has been derived. In this section, we shall derive the failure time distribution for that system both without and with overloading. The method used is an extension of that of Gnedenko [34], which is simple and elegant.

First, we consider a 2-out-of-n system. Let $R(t)$ be the probability that the system is operable up to time t, i.e., 1 - $R(t)$ is a distribution for the system failure, $R(0) = 1$, $R(\infty) = 0$. Since $R(t)$ denotes the probability that the system is operable up to time t, there are three possibilities, viz., (1) all n units are operable up to time t; (2) one unit fails up to time $u\ (0 \leq u < t)$, and the remaining units are operable, but the repair is not completed during the balance of time $t - u$; and (3) one unit fails up to time $u\ (0 \leq u < t)$, the repair is completed during a time interval $v\ (0 \leq u + v < t)$, and consequently the system returns to state 1. These three events are mutually exclusive. Thus we have

$$(2.12)\quad R(t) = e^{-\Lambda t} + \sum_i \frac{\lambda_i}{\Lambda} \int_0^t \Lambda e^{-\Lambda u} e^{-\lambda_i^*(t-u)} \bar{G}_i(t-u)\, du + \sum_i \frac{\lambda_i}{\Lambda} \iint\limits_{u+v<t} \Lambda e^{-\Lambda u} e^{-\lambda_i^* v} dG_i(v)\, R(t-u-v)\, du.$$

This is an integral equation of the renewal type, and the derivation resembles that of the integral equation in renewal theory [14, p.54]. To solve (2.12) for $R(t)$, we apply the LS transforms. Let

$$(2.13)\quad \varphi(s) \equiv \int_0^\infty e^{-st}\, d[1-R(t)] = -\int_0^\infty e^{-st} dR(t)$$

by the LS transform of $1 - R(t)$. Applying the LS transforms to (12), and solving for $\varphi(s)$, we have (see Appendix)

$$(2.14)\quad \varphi(s) = \frac{\sum_i \frac{\lambda_i \lambda_i^*}{s+\lambda_i^*} [1 - g_i(s+\lambda_i^*)]}{s + \sum_i \lambda_i [1 - g_i(s+\lambda_i^*)]},$$

which is the LS transform of the failure time distribution for a 2-out-of-n system.

The MTSF for the system is

$$(2.15)\qquad \mathrm{MTSF} = -\left.\frac{d\varphi(s)}{ds}\right|_{s=0} = \frac{1+\sum_i \lambda_i\,[1-g_i(\lambda_i^*)]/\lambda_i^*}{\sum_i \lambda_i\,[1-g_i(\lambda_i^*)]},$$

which is the same as (2.8).

For a 2-out-of-n system under the overloading condition, we have the following similar integral equation:

$$(2.16)\qquad R(t) = e^{-\Lambda t} + \sum_i \frac{\lambda_i}{\Lambda}\int_0^t \Lambda e^{-\Lambda t} e^{-\lambda_i^*(t-u)}\bar{G}_i(t-u)\bar{A}_i(t-u)\,du$$

$$+\sum_i \frac{\lambda_i}{\Lambda}\iint_{u+v<t} \Lambda e^{-\Lambda u} e^{-\lambda_i^* v}\bar{A}_i(v)\,dG_i(v)\,R(t-u-v)\,du.$$

Applying the LS transforms to (2.16), and solving for $\varphi(s)$, we have (see Appendix)

$$(2.17)\qquad \varphi(s) = \frac{\sum_i \lambda_i\,[A_i(0)+\lambda_i^*\int_0^\infty e^{-(s+\lambda_i^*)t}\bar{G}_i(t)\bar{A}_i(t)\,dt+\int_0^\infty e^{-(s+\lambda_i^*)t}\bar{G}_i(t)\,dA_i(t)]}{s+\sum_i \lambda_i\,[1-\int_0^\infty e^{-(s+\lambda_i^*)t}\bar{A}_i(t)\,dG_i(t)]},$$

which is the LS transform of the failure time distribution for a 2-out-of-n system under the overloading condition.

The MTSF for (2.17), derived as in (2.15), is the same as (2.11).

2. 4. Special Cases

In this section we show that the results of the preceding two sections include many earlier results as special cases.

A. Consider $n = 2$. Then $\lambda_1^* \equiv \lambda_2$, $\lambda_2^* \equiv \lambda_1$. For this case, $\varphi(s)$ in (2.14) and MTSF (T_0) in (2.15) become

$$(2.18)\quad \varphi(s) = \frac{\lambda_1\lambda_2\left[\dfrac{1-g_1(s+\lambda_2)}{s+\lambda_2} + \dfrac{1-g_2(s+\lambda_1)}{s+\lambda_1}\right]}{s+\lambda_1[1-g_1(s+\lambda_2)] + \lambda_2[1-g_2(s+\lambda_1)]},$$

and

$$(2.19)\quad \text{MTSF} = \frac{1+\dfrac{\lambda_1}{\lambda_2}[1-g_1(\lambda_2)] + \dfrac{\lambda_2}{\lambda_1}[1-g_2(\lambda_1)]}{\lambda_1[1-g_1(\lambda_2)] + \lambda_2[1-g_2(\lambda_1)]},$$

which correspond to (4) and (5), respectively, of Gaver [33]. This case is a two-unit paralleled redundant system of dissimilar units (see (1.12) and (1.14) in Section 1.3).

B. Consider the case in which the units are identical. Then $\lambda \equiv \lambda_i$, $G(t) \equiv G_i(t)$ and $\lambda_i^* = (n - 1)$. Then (2.14) and (2.15) become

$$(2.20) \quad \varphi(s) = \frac{n\lambda}{s+n\lambda} \cdot \frac{(n-1)\lambda}{s+(n-1)\lambda} \cdot \frac{1 - g(s+[n-1]\lambda)}{1 - \frac{n\lambda}{s+n\lambda} g(s+[n-1]\lambda)},$$

and

$$(2.21) \quad \text{MTSF} = \frac{1}{(n-1)\lambda} + \frac{1}{n\lambda\{1-g([n-1]\lambda)\}},$$

which correspond to (5.17a)aand (5.19a) of Downton [23].

C. Consider the special case where $n = 2$ in the above. Then

$$(2.22) \quad \varphi(s) = \frac{2\lambda}{s+2\lambda} \cdot \frac{\lambda}{s+\lambda} \cdot \frac{1 - g(s+\lambda)}{1 - \frac{2\lambda}{s+2\lambda} g(s+\lambda)},$$

and

$$(2.23) \quad \text{MTSF} = \frac{1}{\lambda} + \frac{1}{2\lambda[1-g(\lambda)]},$$

which have been given by Gaver [32] and Liebowitz [49]. This case is a two-unit paralleled redundant system of identical units.

D. Consider the case $n = 2$, and set $\lambda_1 \equiv \lambda$, $\lambda \equiv 0$, $\lambda_1^* = \lambda_2^* = \lambda$, $G(t) \equiv G_i(t)$. Then, (2.14) and (2.15) become

$$\varphi(s) = \frac{\frac{\lambda}{s+\lambda}\,[1 - g(s+\lambda)]}{s + \lambda\,[1 - g(s+\lambda)]}, \tag{2.24}$$

and

$$\text{MTSF} = \frac{1}{\lambda} + \frac{1}{\lambda\,[1 - g(\lambda)]} \tag{2.25}$$

which correspond to (4) and (5), respectively, of Gaver [32]. This case is a two-unit standby redundant system (see Section 1.4).

E. Consider the case $n \rightarrow n + 1$, $\lambda_1 \equiv \lambda_1$, $\lambda \equiv \lambda_j$ ($j = 2, \ldots, n + 1$), $G(t) \equiv G_j(t)$ and $\lambda_j^* = n\lambda$ ($j = 1, \ldots, n + 1$). That is, we consider a system which is composed of n units plus one standby unit. Then (2.14) and (2.15) become

$$\varphi(s) = \frac{n\lambda\,(n\lambda + \lambda_1)\,[1 - g(s+n\lambda)]}{(s + n\lambda)\,\{s + (n\lambda + \lambda_1)\,[1 - g(s+n\lambda)]\}}, \tag{2.26}$$

and

$$\text{MTSF} = \frac{n\lambda + (n\lambda + \lambda_1)\,[1 - g(n\lambda)]}{n\lambda\,(n\lambda + \lambda_1)\,[1 - g(n\lambda)]}. \tag{2.27}$$

In cases D and E, we did not use equation (2.3) ($\lambda_i^* \equiv \Lambda - \lambda_i$) since the calculation for $\varphi(s)$ does not require it.

2. 5. Conclusion

We have discussed a system of dissimilar units. The method of the integral equations for obtaining the failure time distribution is simple and elegant. The generalization to a system under an overloading condition is also simple and may be applied to other modified systems.

Other methods of handling this problem are, e.g., the supplementary variable technique, or the application of semi-Markov processes. But the supplementary variable technique requires that the distribution be absolutely consinuous, and the calculation is complicated. Our method does not require the absolute continuity of the distribution, and is simpler to calculate.

Appendix

In this appendix we show only the derivation of $\varphi(s)$ in (2.17) because $\varphi(s)$ in (2.14) is immediately obtained by setting $A_i(t) \equiv 0$.

Applying the Laplace transforms to (2.16), we have†:

L.H.S. = $[1 - \varphi(s)]/s$.

The first term of R.H.S. = $1/[s + \Lambda]$.

The second term of R.H.S.

$$= \sum \lambda_i \int_0^\infty e^{-st} e^{-\lambda_i^* t} dt \int_0^t e^{-(\Lambda-\lambda_i^*)u} \bar{G}_i(t-u) \bar{A}_i(t-u) du$$

$$= \sum \lambda_i \int_0^\infty e^{-(\Lambda-\lambda_i^*)u} du \int_u^\infty e^{-(s+\lambda_i^*)t} \bar{G}_i(t-u) \bar{A}_i(t-u) du$$

$$= \sum \lambda_i \int_0^\infty e^{-(\Lambda-\lambda_i^*)u} e^{-(s+\lambda_i^*)u} du \int_0^\infty e^{-(s+\lambda_i^*)t} \bar{G}_i(t) \bar{A}_i(t) dt$$

$$= \sum \frac{\lambda_i}{s+\Lambda} \int_0^\infty e^{-(s+\lambda_i^*)t} \bar{G}_i(t) \bar{A}_i(t) dt.$$

The third term of R.H.S.

$$= \sum \lambda_i \int_0^\infty e^{-st} dt \int_0^t e^{-\lambda_i^* v} \bar{A}_i(v) dG_i(v) \int_0^{t-v} e^{-\Lambda u} R(t-u-v) du$$

† R.H.S. and L.H.S. stand for right and left hand side. $\sum \equiv$ sum over i from 1 to n.

$$= \sum \frac{\lambda_i}{\Lambda} \int_0^\infty e^{-st} dt \int_0^t e^{-\lambda_i^* v} \bar{A}_i(v)\, dG_i(v)\, [F(t-v) * R(t-v)]$$

(where $F(t) \equiv 1 - e^{-\Lambda t}$ and $F(t) * R(t)$ denotes the convolution of $F(t)$ and $R(t)$.)

$$= \sum \frac{\lambda_i}{\Lambda} \int_0^\infty e^{-\lambda_i^* v} \bar{A}_i(v)\, dG_i(v) \int_v^\infty e^{-st} [F(t-v) * R(t-v)]\, dt$$

$$= \sum \frac{\lambda_i}{\Lambda} \int_0^\infty e^{-(s+\lambda_i^*)v} \bar{A}_i(v)\, dG_i(v) \int_0^\infty e^{-st} [F(t) * R(t)]\, dt$$

$$= \sum \frac{\lambda_i}{\Lambda} \int_0^\infty e^{-(s+\lambda_i^*)v} \bar{A}_i(v)\, dG_i(v)\, \frac{\Lambda}{s+\Lambda} \cdot \frac{1-\varphi(s)}{s}$$

$$= \sum \frac{\lambda_i}{s+\Lambda} \cdot \frac{1-\varphi(s)}{s} \int_0^\infty e^{-(s+\lambda_i^*)v} \bar{A}_i(v)\, dG_i .$$

Thus, setting L.H.S = R.H.S and solving for $\varphi(s)$, we have (2.17).

CHAPTER III

SIGNAL FLOW GRAPH ANALYSIS FOR SYSTEMS

3. 1. Introduction

Markov chains are well-known as a mathematical tool for system analysis. Renewal processes are also used to analyze sytems (in particular, maintainable systems). A Markov renewal process (or semi-Markov process), which is a marriage of Markov chains and renewal processes, was first discussed by Lévy and Smith, independently, in 1954. A Markov renewal process is one of the most important mathematical tools for system analysis. We shall discuss Markov renewal processes as a mathematical tool throughout this chapter.

Graphical representations for systems are of great importance in system science. Especially, block diagrams, signal flow graphs, and wiring diagrams are generally used to represent systems graphically. Block diagrams are used for control engineering, signal flow graphs are used for electrical engineering (in particular, electrical circuit theory), and wiring diagrams are used for analogue computation (or simulation). The relationships among the above three graphs (or diagrams) are well-known (see, e.g., Huggins [42, 43]). Here we shall discuss the relationship between Markov renewal processes and signal flow graphs, and show that deriving the Laplace-Stieltjes (LS) transform of the first passage time distribution from one state to another in a Markov renewal process is obtaining the system gain by defining that a starting state is a source and an ending state is a sink in the signal flow graph.

Using the relationship between Markov renewal processes and signal flow graphs, we shall finally obtain system reliability for some systems, e.g., a two-unit standby redundant system, a two-unit standby redundant system with noninstantaneous switchover, and m-out-of-n systems. The use of signal flow graphs for system analysis makes the system clear, and obtaining the system gain implies our desired result, which is an easy mechanical procedure.

3. 2. Markov Renewal Processes

A Markov renewal process [53, 54], roughly speaking, is a stochastic process in which the state transitions obey the given transition probabilities and the sojourn time in a state is a random variable with any distribution depending on that state and the next visiting state, where the number of states may be denumerable. In this paper we restrict our attention to Markov renewal processes with finitely many states since our models can be usually represented by Markov renewal processes with finitely many states. The detailed discussion of Markov renewal processes with finitlly many states can be found in Pyke [54].

Here we shall describe the necessary definitions and properties of these processes. We denote the states of a Markov renewal process by the symbols S_0, S_1, ..., S_N. We define the transition probability p_{ij} from state S_i to state S_j for all

$i, j = 0, 1, 2, \ldots, N$. We also define the distribution $F_{ij}(t)$ ($t \geq 0$) of the sojourn time in state S_i and the next visiting state S_j. We define

$$(3.1) \qquad Q_{ij}(t) = p_{ij} F_{ij}(t). \qquad (i, j = 0, 1, \ldots, N)$$

Defining the states of the process, we can find the $Q_{ij}(t)$ for all i and j. Then we have all information on the process considered. In this chapter we restrict our attention to the first passage time distribution from one state to another state. So, define an absorbing state S_N. Then the remaining states $i = 0, 1, 2, \ldots, N - 1$ are transient. We define the first passage time distribution $\Phi_i(t)$ ($i = 0, 1, 2, \ldots, N - 1$) starting from state S_i at $t = 0$ to the absorbing state S_N up to time t. We can consider two cases for the $\Phi_i(t)$: One is the immediate transition to the absorbing state S_N. Another is the transition to any transient state S_j ($j = 0, 1, \ldots, N - 1$). These two events are mutually exclusive. In the latter case, the process after the transition to state S_j obeys the $\Phi_j(t)$. Thus we have

$$(3.2) \qquad \Phi_i(t) = Q_{iN}(t) + \sum_{j=0}^{N-1} Q_{ij}(t) * \Phi_j(t),$$

$$(i = 0, 1, 2, \ldots, N - 1)$$

where $*$ denotes the convolution. The small letters $\varphi_i(s)$ and $q_{ij}(s)$ denote the corresponding LS transforms of the $\Phi_i(t)$ and $Q_{ij}(t)$, respectively. Defining the $N \times 1$ vector $\varphi(s)$ with component $\varphi_i(s)$ ($i = 0, 1, \ldots, N-1$) and taking the LS transforms for (3.2), we have in matrix form

$$\varphi(s) = g_N(s) + g(s)\varphi(s), \tag{3.3}$$

where the $g_N(s)$ is the $N \times 1$ vector with component $g_{iN}(s)$ ($i = 0, 1, \ldots, N-1$) and the $g(s)$ is the $N \times N$ matrix with element $g_{ij}(s)$ ($i, j = 0, 1, \ldots, N-1$). Solving (3.3) for $\varphi(s)$, we have

$$\varphi(s) = [I - g(s)]^{-1} g_N(s), \tag{3.4}$$

where I is the $N \times N$ identity matrix. Our concern is to find the $\varphi(s)$ (in particular, $\varphi_0(s)$) for the models discussed below. The general first passage time distribution from one state to another has been given by Pyke [54]. Using the general result, we can also obtain (3.4). We, however, derived (3.4) by using the intuitive method.

3. 3. Signal Flow Graphs

In this section we shall consider an algorithm for deriving the $\varphi_i(s)$ ($i = 0, 1, 2, \ldots, N-1$) by using signal flow graphs. The definitions and notations of signal flow graphs obey those of Chow and Cassignal [13].

Consider a system whose states are defined and their associated $\gamma_{ij}(s)$'s are given. For the system, using the states and the $\gamma_{ij}(s)$, we can construct a state transition diagram which becomes a signal flow graph of the system. In the graph each node corresponds to each state of the system and each branch gain to the $\gamma_{ij}(s)$. We shall consider an algorithm for deriving the $\varphi_0(s)$ by using the signal flow graph. As is anticipated, the $\varphi_0(s)$ given in (3.4) is derived by using Mason's gain formula, [13, p.63] in the signal flow graph, where we define that node S_0 is a source and node S_N is a sink. We shall below verify the above fact by using the results of signal flow graphs [13].

Since the source S_0 in the graph has both incoming and outcoming branches, we define a new source S_a which has an outcoming branch to node S_0 with its branch gain unity (see Fig. 3.1). From equation (A.5) in reference 13, p.136, we have

$$(3.5) \qquad T_{Na} \equiv \frac{x_N}{x_a} = -\frac{\Delta_{ON}}{\Delta},$$

where the $(N+1) \times (N+1)$ determinant

$$(3.6) \qquad \Delta \equiv \begin{vmatrix} a_{00} & a_{01} & \cdots & a_{0N} \\ a_{10} & a_{11} & \cdots & a_{1N} \\ \vdots & \vdots & & \vdots \\ a_{N0} & a_{N1} & \cdots & a_{NN} \end{vmatrix},$$

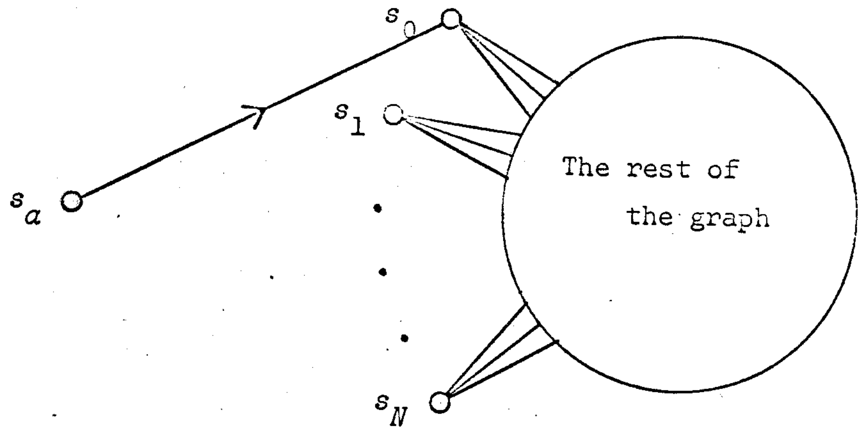

Fig. 3.1. Signal flow graph with the source connected only to one node of the system.

and the Δ_{0N} is the (0, N) cofactor of the Δ. Each element a_{ji} of the Δ can be represented by

$$(3.7) \qquad a_{ji} = \delta_{ij} - g_{ij}(s).$$

$$(i, j = 0, 1, 2, \ldots, N)$$

Here δ_{ij} denotes the Kronecker delta. We note that the subscripts i and j in the a_{ji} and $g_{ij}(s)$ are interchanged. In the system considered, we have that $g_{Nj}(s) = 0$ ($j = 0, 1, 2, \ldots, N - 1$) and $g_{NN}(s) = 1$ since node S_N is absorbing. If we set $g_{NN}(S) = 1$, the determinant Δ is singular. So we set $g_{NN}(S) = 0$ ($a_{NN} = 1$) for the convenience of analysis. Defining the $N \times N$ determinant

$$\Delta' = \begin{vmatrix} a_{00} & \cdots & a_{0,N-1} \\ \vdots & & \vdots \\ a_{N-1,0} & \cdots & a_{N-1,N-1} \end{vmatrix}, \tag{3.8}$$

we have $\Delta = \Delta'$ from the above fact. We also note that $\Delta' = \left| I - \mathfrak{g}(s)^T \right| = \left| I - \mathfrak{g}(s) \right|$, where the superscript T denotes the transpose of the matrix and the $\mathfrak{g}(s)$ is defined in (3.3). Then we have from (3.5)

$$T_{Na} = -\Delta_{0N} / \Delta'. \tag{3.9}$$

Expanding the cofactor Δ_{0N} with respect to the th row and using $a_{Nj} = -g_{jN}(s)$ ($j = 0, 1, 2, \ldots, N-1$), we have

$$\begin{aligned} \Delta_{0N} &= \sum_{j=0}^{N-1} a_{Nj} \Delta'_{0j} \\ &= -\sum_{j=0}^{N-1} g_{jN}(s) \Delta'_{0j}, \end{aligned} \tag{3.10}$$

where the Δ'_{0j} is the $(0, j)$ cofactor of the Δ'. Thus we have

$$(3.11) \quad T_{N\alpha} = -\Delta_{0N} / \Delta'$$

$$= \sum_{j=0}^{N-1} g_{jN}(s) \Delta'_{0j} / \Delta'$$

$$= \{ g_N(s)^T [I - g(s)^T]^{-1} \}_{j=0}$$

$$= \{ [I - g(s)]^{-1} g_N(s) \}_{i=0} ,$$

where $\{ \ \}_i$ denotes the ith component of the vector. Therefore, we show that equation (3.11) coincides with (3.4). For the other $\varphi_i(s)$ ($i = 1, 2, \ldots, N - 1$), we can show the same result. In the above discussion we defined a new source S_α. In practice, we define that a starting state is a source and an ending state is a sink, and by using Mason's gain formula we can obtain the system gain which is the LS transform of the first passage time distribution.

3. 4. System Reliability

In the preceding section we discussed the relationship between Markov renewal processes and signal flow graphs.

Markov renewal processes are of great use for system analysis. In particular, the processes are used in the reliability theory. We encounter a problem that the total failure of a system yields a catastrophe. Our concern in the problem is the first passage time distribution to system down. For the problem we can apply the signal flow graph method to obtain the LS transform of the first passage time distribution. We shall show some examples of systems. The systems have been investigated by Gaver [32, 33], Srinivasan [58, 60], Downton[23], and others. However, the signal flow graph method discussed in this chapter is simple and elegant.

A Two-Unit Standby Redundant Systems

A two-unit standby redundant system with instantaneous switchover has been investigated by Gnedenko et al. [35], and Srinivasan [60] under the most generalized assumption that both the failure and repair time distributions are arbitrary. Appropriately labeling the number of the two units, we may call them units 1 and 2. The failure time of unit i (i = 1, 2) is a random variable with an arbitrary distribution $F_i(t)$ and the repair time of unit i is also a random variables with an arbitrary distribution $G_i(t)$. These random variables are nonnegative and mutually independent.

Initially at t = 0, unit 1 begins to be operating and unit 2 is in standby (state S_0). As soon as unit 1 fails, unit 2 begins to be operating and unit 1 undergoes repair (state S_1). When the repair of unit 1 is completed before unit 2 fails, unit 1 is in standby and then as soon as unit 2 fails, unit 1 in standby begins to be operating and unit 2 undergoes repair (state S_2). While in state S_1 we consider another case that unit 2 fails before the repair completion of unit 1, which implies the system down (state S_3). In state

we can consider

two cases: One is the repair completion of unit 2 before unit 1 fails, which goes to state S_1. Another is the failure of unit 1 before the repair completion of unit 2, which goes to state S_3. The system behaves from the operating unit 2 to the operating unit 1 and so on until the recurrence of the system down. In the system considered, we assume that each switchover time is instantaneous.

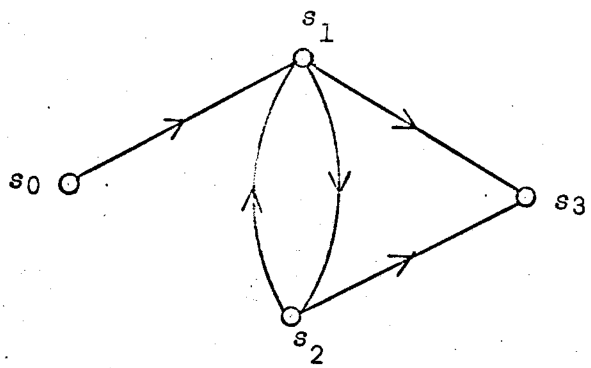

Fig. 3.2. Signal flow graph of a two-unit standby redundant system.

Fig. 3.2 shows the signal flow graph of the system. Then each branch gain is easily obtained from the discussion just mentioned above as follows:

(3.12) $$g_{01}(s) = \int_0^\infty e^{-st}\, dF_1(t) \equiv f_1(s),$$

(3.13) $$g_{12}(s) = \int_0^\infty e^{-st}\, G_1(t)\, dF_2(t),$$

$$(3.14) \qquad g_{21}(s) = \int_0^\infty e^{-st} G_2(t)\,dF_1(t),$$

$$(3.15) \qquad g_{13}(s) = \int_0^\infty e^{-st} \bar{G}_1(t)\,dF_2(t),$$

$$(3.16) \qquad g_{23}(s) = \int_0^\infty e^{-st} \bar{G}_2(t)\,dF_1(t),$$

where $\bar{G}_i(t) = 1 - G_i(t)$ is the survival probability function. In the graph in Fig. 3.2, node S_0 is a source and node S_3 is a sink. Thus, we have from Mason's gain formula [13]the following system gain

$$(3.17) \qquad \varphi_0(s) = \frac{g_{01}(s)\, g_{12}(s)\, g_{23}(s) + g_{01}(s)\, g_{13}(s)}{1 - g_{12}(s)\, g_{21}(s)},$$

which is the LS transform of the first passage time distribution from state S_0 to state S_3. The mean time is given by

$$(3.18) \qquad \hat{T} = -\left.\frac{d\varphi_0(s)}{ds}\right|_{s=0}$$

$$= \frac{1}{\lambda_1} + \frac{1/\lambda_2 + g_{12}(0)/\lambda_1}{1 - g_{12}(0)\, g_{21}(0)},$$

where

$$(3.19) \qquad 1/\lambda_i = \int_0^\infty t\,dF_i(t). \qquad (i = 1, 2)$$

A Two-Unit Standby Redundant System with Noninstantaneous Switchover

Here we shall discuss a two-unit standby redundant system with noninstantaneous switchover, which has been discussed by Srinivasan [60]. He discussed a simple case that two units are identical, but we shall discuss the more generalized system that two units are dissimilar. In the similar way of the preceding system, the two units are denoted by i = 1, 2. The failure time of unit i is a random variable with distribution $F_i(t) = 1 - \exp(-\lambda_i t)$ and the repair time of unit i is a random variable with an arbitrary distribution $G_i(t)$. Here we assume the memoryless property of the failure time distribution. Whenever unit i (i = 1, 2) is active and the other j ($j \neq i$) in standby, action is initiated on the latter of T_j unit of time in order to bring it to the operating standby state. The switchover time from the action to the operating standby state of unit j is a random variable with an arbitrary distribution $\Gamma_j(t)$. These random variables are nonnegative and mutually independent. The five states of each unit are active, repair, standby, switchover, and operating standby, and are denoted by the symbols 0, 1, 2, 3, and 4, respectively. The state of the system will be specified by the states of unit 1 and unit 2 together. The possible state of the system are enumerated in Table 3.1, where states S_8, S_9, S_{10}, S_{11}, and S_{12} denote the system down and these states

Table 3. 1. Possible states of the system.

State of system	s_0	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}
State of unit 1	0	0	0	1	2	3	4	0	1	1	1	3	2
State of unit 2	2	3	4	0	0	0	0	1	1	2	3	1	1

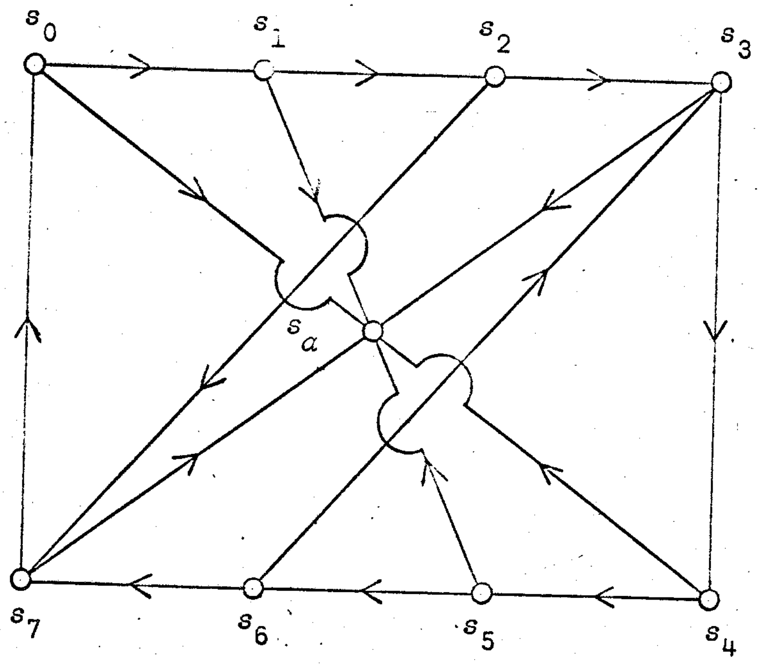

Fig. 3. 3. Signal flow graph of a two-unit standby redundant system with noninstantaneous switchover.

are combined in an absorbing state S_a.

Fig. 3.3 shows the signal flow graph of the system using state $S_0 - S_7$ and S_a. Each branch gain of the system is given by

(3.20) $$g_{01}(s) = e^{-(s+\lambda_1)T_2}, \qquad g_{45}(s) = e^{-(s+\lambda_2)T_1}$$

(3.21) $$g_{0a}(s) = \frac{\lambda_1}{s+\lambda_1}\left[1 - e^{-(s+\lambda_1)T_2}\right], \quad g_{4a}(s) = \frac{\lambda_2}{s+\lambda_2}\left[1 - e^{-(s+\lambda_2)T_1}\right]$$

(3.22) $$g_{12}(s) = \gamma_2(s+\lambda_1), \qquad g_{56}(s) = \gamma_1(s+\lambda_2)$$

(3.23) $$g_{1a}(s) = \frac{\lambda_1}{s+\lambda_1}\left[1 - \gamma_2(s+\lambda_1)\right], \quad g_{5a}(s) = \frac{\lambda_2}{s+\lambda_2}\left[1 - \gamma_1(s+\lambda_2)\right]$$

(3.24) $$g_{23}(s) = \frac{\lambda_1}{s+\lambda_1+\lambda_2}, \qquad g_{67}(s) = \frac{\lambda_2}{s+\lambda_1+\lambda_2}$$

(3.25) $$g_{27}(s) = g_{67}(s), \qquad g_{63}(s) = g_{23}(s)$$

(3.26) $$g_{34}(s) = g_1(s+\lambda_2), \qquad g_{70}(s) = g_2(s+\lambda_1)$$

(3.27) $$g_{3a}(s) = \frac{\lambda_2}{s+\lambda_2}\left[1 - g_1(s+\lambda_2)\right], \quad g_{7a}(s) = \frac{\lambda_1}{s+\lambda_1}\left[1 - g_2(s+\lambda_1)\right],$$

where $g_i(s)$ and $\gamma_i(s)$ (i = 1, 2) are the LS transforms of $G_i(t)$ and $\Gamma_i(t)$, respectively. We define that node S_0 is a source and S_a is a sink in the system. From Mason's gain formula the system gain is given by

$$\varphi_0(s) = N/(1 - g_{01}g_{12}g_{27}g_{70} - g_{45}g_{56}g_{63}g_{34}), \tag{3.28}$$

where

$$N = g_{0a} + g_{01}g_{1a} + g_{01}g_{12}g_{23}g_{3a} + g_{01}g_{12}g_{27}g_{7a} + g_{01}g_{12}g_{23}g_{34}g_{4a} - g_{0a}g_{45}g_{56}g_{63}g_{34} + g_{01}g_{12}g_{23}g_{34}g_{45}g_{5a} - g_{01}g_{1a}g_{45}g_{56}g_{63}g_{34}. \tag{3.29}$$

Here we use the abbreviated notation g_{ij} instead of $g_{ij}(s)$. We further note that the results (3.28) and (3.29) are simplified by using the relations (3.20)-(3.27). We have obtained the $\varphi_i(s)$, which is the LS transform of the first passage time distribution to system down from state S_0. We can obtain the $\varphi_i(s)$ (i = 1, ..., 7) by defining a source S_i in the similar fashion.

Srinivasan [60] considered a special case of the above system. That is, he considered a simple case that two units are identical. In this case we assume that the failure time distribution is $F(t) = 1 - \exp(-\lambda t)$, the repair time distribution $G(t)$, the switchover time distribution $\Gamma(t)$, and the requited time to bring a standby unit to the operating standby state T. Then each branch gain is given in (3.20)-(3.27), where the two branch gains for the right and left equations are identical. Noting that states S_0 and S_4, S_1 and S_5, S_2 and S_6, and S_3 and S_7 are identical, we have the reduced signal flow graph in Fig. 3.4.

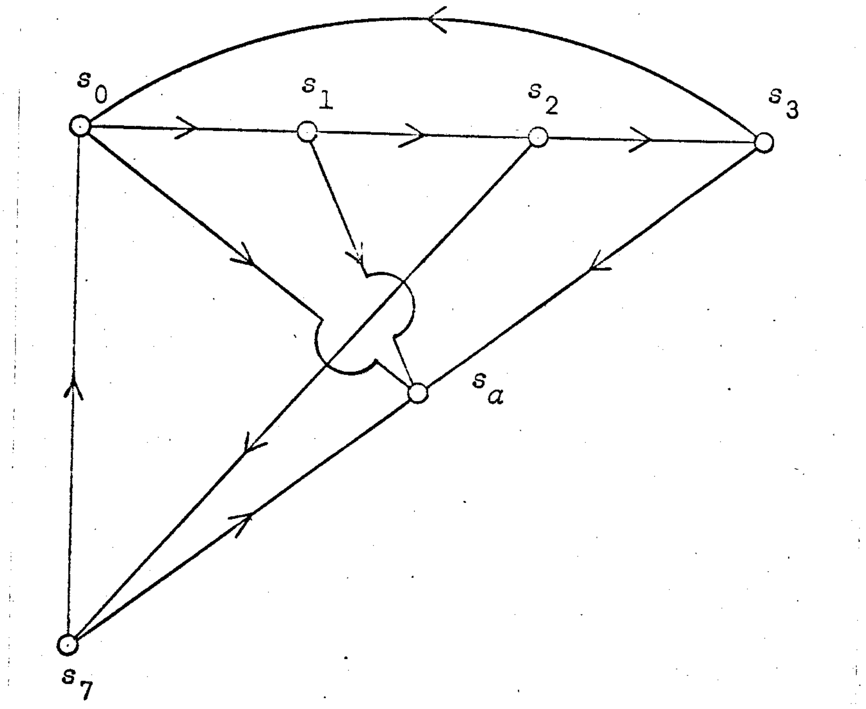

Fig. 3.4. Signal flow graph of the system of identical units.

Assuming that node S_0 is a source and node S_a is a sink, we have the system gain from Mason's gain formula as follows:

$$(3.30) \qquad \varphi_0(s) = \frac{q_{0a} + q_{01}q_{1a} + 2\,q_{01}q_{12}q_{23}q_{3a}}{1 - 2\,q_{01}q_{12}q_{27}q_{70}}.$$

The mean time is given by

$$(3.31) \qquad \hat{T} = \frac{1}{\lambda}\left\{1 + \frac{1}{2\,[1 - r(\lambda)g(\lambda)e^{-\lambda T}]}\, r(\lambda)\, e^{-\lambda T}\right\}.$$

m-out-of-n System

An m-out-of-n system is a redundant system composed of n paralleled units $(n \geq m)$. When m units are simaltaneously under failure or repair, the system down occurs [23]. We assume that the system considered has one repair facility and the failed unit may be waiting for the repair if the repair facility is busy. We also assume that each switchover time is instantaneous.

First, we shall consider a 2-out-of-n system of dissimilar units. Appropriately labeling the number of units, we may call them units 1, 2, ..., n - 1, and n. The failure time of unit i (i = 1, 2, ..., n) is a random variable with exponential distribution $F_i(t) = 1 - \exp(-\lambda_i t)$ and the repair time of the failed unit i is a random variable with an arbitrary distribution $G_i(t)$, where we assume the memoryless property of failure time distribution for the convenience of analysis. These random variables are nonnegative and mutually independent. In the system state S_0 denotes one that all n units are operating, state S_i ($i = 1, 2, \ldots, n$) denotes one that unit i is under repair and the remaining units are operating, and state S_{n+1} denotes one that at least two units are under repair or failure (i.e., state S_{n+1} denotes the system down). Fig. 3.5 shows the signal flow graph of the system. For the system

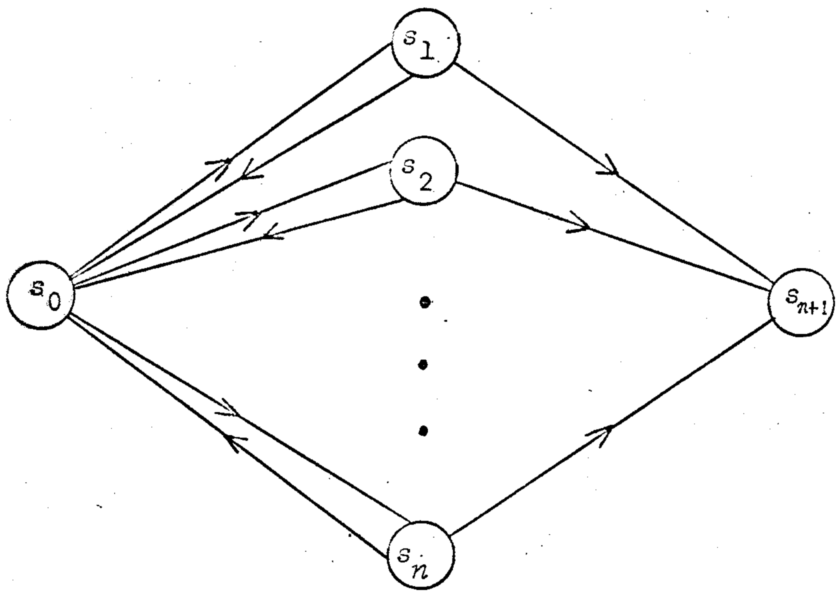

Fig. 3.5. Signal flow graph of a 2-out-of-n system of dissimilar units.

we obtain easily each branch gain as follows:

$$(3.32) \qquad g_{0i}(s) = \lambda_i \,/\, (s+\Lambda), \qquad (i = 1, 2, \ldots, n)$$

(3.33) $$g_{i0}(s)=\int_0^{\infty} e^{-st}e^{-\lambda_i^* t}\,dG_i(t)$$

$$=g_i(s+\lambda_i^*), \qquad (i = 1, 2, \ldots, n)$$

(3.34) $$g_{i,n+1}(s)=\int_0^{\infty} e^{-st}\bar{G}_i(t)\lambda_i^* e^{-\lambda_i^* t}\,dt$$

$$=\frac{\lambda_i^*}{s+\lambda_i^*}\left[1-g_i(s+\lambda_i^*)\right],$$

$$(i = 1, 2, \ldots, n)$$

where

(3.35) $$\Lambda=\sum_{i=1}^{n}\lambda_i ,$$

(3.36) $$\lambda_i^*=\Lambda-\lambda_i . \qquad (i = 1, 2, \ldots, n)$$

Assuming that node S_0 is a source and node S_{n+1} is a sink, we have from Mason's gain formula

(3.37) $$\varphi_0(s)=\sum_{i=1}^{n} g_{0i}(s)\,g_{i,n+1}(s)\Big/\left[1-\sum_{i=1}^{n} g_{0i}(s)\,g_{i0}(s)\right].$$

The mean time can be immediately obtained from (3.37), but we omit the result (see Section 2.3). As a special case of the above system, we consider $n = 2$. In this case the system is a two-unit paralleled redundant system, which is used in many fields.

Second, we shall consider a 3-out-of-n system.

In the system,we shall only consider a simple case that all units are identical. The failure time distribution of each unit is $F(t) = 1 - \exp(-\lambda t)$ and the repair time distribution of each unit is $G(t)$. The state S_i (i = 0, 1, 3) of the system denotes the corresponding number of the failed units. In the system we need not to consider a state S_2 since we take notice of the regeneration point of the repair

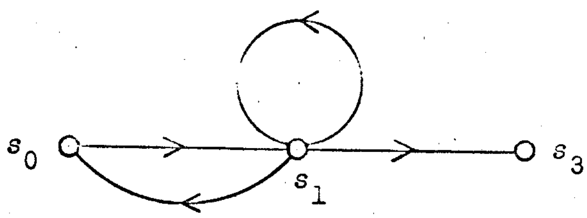

Fig. 3.6. Signal flow graph of a 3-out-of-n system of identical units.

time distribution. Fig. 3.6 shows the signal flow graph of the system. Each branch gain is given by

(3.38) $$g_{01}(s) = n\lambda / (s + n\lambda),$$

(3.39) $$g_{10}(s) = \int_0^\infty e^{-st} e^{-(n-1)\lambda t} dG(t) = g(s + [n-1]\lambda),$$

(3.40) $$g_{11}(s) = \binom{n-1}{1} \int_0^\infty e^{-st} (1 - e^{-\lambda t}) e^{-(n-2)\lambda t} dG(t)$$

$$= (n-1) \{ g(s + [n-2]\lambda) - g(s + [n-1]\lambda) \},$$

$$(3.41) \quad g_{13}(s) = \binom{n-1}{2} \int_0^\infty e^{-st}\, \bar{G}(t)\, e^{-(n-3)\lambda t} d[(1-e^{-\lambda t})^2]$$

$$= \frac{\lambda(n-1)(n-2)}{s+(n-2)\lambda}[1-g(s+[n-2]\lambda)] - \frac{\lambda(n-1)(n-2)}{s+(n-1)\lambda}[1-g(s+[n-1]\lambda)].$$

Assuming that node S_0 is a source and node S_3 is a sink, we have from Mason's gain formula

$$(3.42) \qquad \varphi_0(s) = \frac{g_{01}(s)\, g_{13}(s)}{1 - g_{01}(s)\, g_{10}(s) - g_{11}(s)} \quad .$$

The mean time can be immediately obtained from (3.42)

As the third model we shall consider a 4-out-of-n system. In the system we shall also consider a simple case that all units are identical. The state S_i (i = 0, 1, 2, 4) of the system denotes the corresponding number of the failed units. In the same reason of the preceding model we need not to consider a state S_3. Fig. 3.7 shows the signal flow graph of

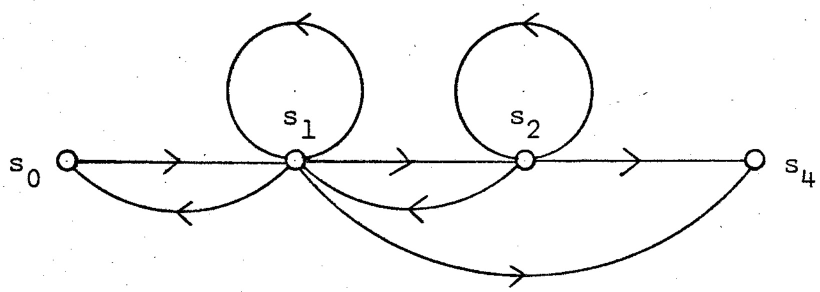

Fig. 3.7. Signal flow graph of a 4-out-of-n system of identical units.

the system. We shall derive each branch gain. For $g_{01}(s)$, $g_{10}(s)$, and $g_{12}(s)$, we have obtained in (3.38), (3.39), and (3.40), respectively, of a 3-out-of-n system. For the other $g_{ij}(s)$, we have by focussing on the regeneration point of the repair time distribution

$$(3.43) \quad g_{12}(s) = \frac{(n-1)(n-2)}{2}\left[g(s+[n-3]\lambda) - 2g(s+[n-2]\lambda) + g(s+[n-1]\lambda)\right],$$

$$(3.44) \quad g_{14}(s) = \frac{\lambda(n-1)(n-2)}{s+(n-2)\lambda}\left[1-g(s+[n-2]\lambda)\right] - \frac{\lambda(n-1)(n-2)}{s+(n-1)\lambda}\left[1-g(s+[n-1]\lambda)\right],$$

$$(3.45) \quad g_{21}(s) = g(s+[n-2]\lambda),$$

$$(3.46) \quad g_{22}(s) = (n-2)\left[g(s+[n-3]\lambda) - g(s+[n-2]\lambda)\right],$$

$$(3.47) \quad g_{24}(s) = \frac{(n-2)\lambda}{s+(n-2)\lambda}\left[1-g(s+[n-2]\lambda)\right].$$

Assuming that node S_0 is a source and node S_4 is a sink, we have from Mason's gain formula

$$(3.48) \quad \varphi_0(s) = \frac{g_{01}g_{12}g_{24} + g_{01}g_{14}(1-g_{22})}{1-g_{01}g_{10}-g_{11}-g_{12}g_{21}-g_{22}+g_{01}g_{10}g_{22}+g_{11}g_{22}},$$

where we also use the abbreviated notation g_{ij} instead of $g_{ij}(s)$. The mean time can be immediately obtained from (3.48).

In general, we can consider an m-out-of-n system, where we assume that all units are identical. The similar signal flow graph can be demonstrated and each branch gain can be obtained. Thus we can obtain the LS transform of the first passage time distribution to system down for an m-out-of-n system. Here we omit the detailed results.

3. 5. Conclusion

The relationship between continuous time Markov processes and signal flow graphs have been discussed by Tin Htun [62] and Dolazza [22]. So far as we know, we have found no paper describing the relationship between Markov renewal processes and signal flow graphs. Markov renewal processes are of great use for the analysis in system science since the processes are generalizations of Markov processes and renewal processes, and have the fruitful results [53, 54]. We believe that the results obtained in this chapter are of great use and may be applicable to many other fields.

The signal flow graph approach is intuitive and obtaining the required quantity is an easy mechanical procedure form Mason's gain formula. Thus, the LS transforms of the first passage time distribution can be automatically obtained if we can give the signal flow graph and its associated branch gains.

CHAPTER IV

A TWO-UNIT STANDBY REDUNDANT SYSTEM WITH STANDBY FAILURE

4. 1. Introduction

We have discussed some redundant repairable models in the preceding chapters. In this chapter we shall also consider a two-unit standby redundant system. The analysis for the model in Sections 1. 4 and 3. 4 has been made with the assumption that a standby unit never fails in the interval of standby state. In this chapter we should, however, consider the failure of standby units in that interval.

In the first analysis for the model we shall discuss the system with exponential failure, general repair, and general standby failure. Then we shall give the Laplace-Stieltjes (LS) transform of each transition time distribution from one state to another, where a state is the time instant of the system. Using the LS transform of each transition time distribution (which is a branch gain of the signal flow graph considered), and applying Mason's gain formula, we shall derive the LS transform of the time distribution to first system down. As special cases we shall discuss a two-unit paralleled redundant system with exponential failure and general repair, and a two-unit standby redundant system with exponential failure and general repair.

In the second analysis for the model we shall discuss the system with all general failure, repair, and standby failure. Focussing on the regeneration point of the failure time distribution, we shall derive the LS transform of the time distribution to first system down. In the analysis we cannot obtain all one step transition time distributions. The similar special cases of the first analysis will be discussed.

4. 2. Model with Exponential Failure

We consider a two-unit standby redundant system of two dissimilar units, which is considered to be the most generalized model. Appropriately labelling the number of two units, we may call unit i (i = 1, 2). The failure time of the operative unit i is subject to the exponential distribution $F_i(t) = 1 - \exp(-\lambda_i t)$ ($t \geq 0$), the repair time of unit i is subject to an arbitrary distribution $G_i(t)$ ($t \geq 0$), and the failure time of the standby unit i in the standby interval is subject to an arbitrary distribution $H_i(t)$ ($t \geq 0$). Here we assume that the failure time of an operative unit has the "memoryless property," i.e., the failure time distribution is exponential. We use the memoryless property for the analysis as it is shown below. We shall, however, show the general failure case in Section 4.5. We define the survival distribution $\bar{H}_i(t) \equiv 1 - H_i(t)$, which denotes the probability that the standby unit i does not fail up to time t in the standby interval. We also define the similar survival distributions $\bar{F}_i(t) \equiv 1 - F_i(t)$ and $\bar{G}_i(t) \equiv 1 - G_i(t)$.

We assume that the switchover times from the operative state to the repair, from the repair completion to the standby state, and from the standby state to the operative state are instantaneous. We further assume that the repair completion of a unit recovers its functioning perfectly. We note that, when the standby unit i is put into operation, the failute time of the operative unit i also obeys $F_i(t)$, which is independent of the standby time.

4. 3. Derivation of the LS Transform

We shall apply the signal flow graph method of analysing the model just mentioned above. Our concern is the time distribution to first system down, where we assume that initially at $t = 0$ unit 1 begins to be operative and at that time unit 2 begins to be standby.

In our model, we define the following five states, where a state is the time instant (or the epoch) of the system.

State S_0; unit 1 begins to be operative and unit 2 begins to be standby.

State S_1; unit 1 begins to be standby and unit 2 begins to be operative.

State S_2; unit 1 begins to get repaired and unit 2 begins to be operative.

State S_3; unit 1 begins to be operative and unit 2 begins to get repaired.

State S_4; two units are under repair or failure, which denotes the system down.

Using the above states, we obtain the signal flow graph of the system in Fig. 4.1. We shall give each branch gain.

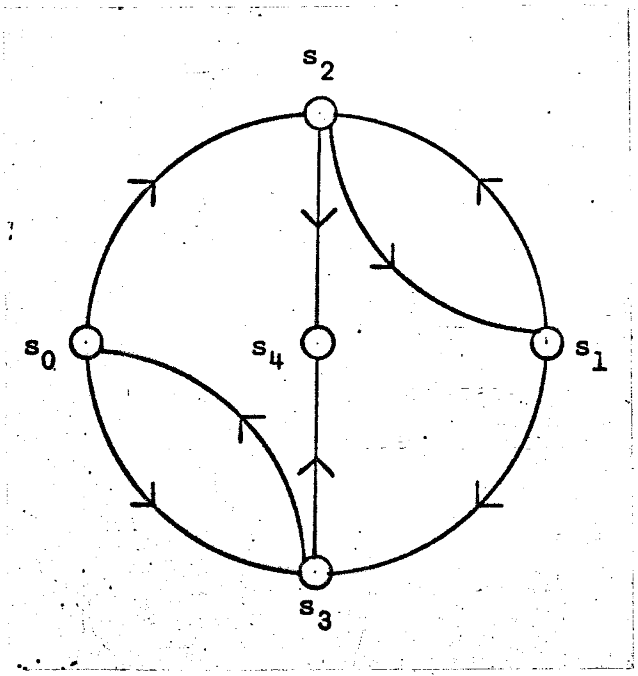

Fig. 4.1. Signal flow graph of the system.

In state S_0 we can consider the following two (exclusive and exhaustive) cases:

(i) The operative unit 1 fails before the standby unit 2 fails.

(ii) The standby unit 2 fails before the operative unit 1 fails.

In the first case the system goes to state S_2. Its branch gain is

$$(4.1)\qquad g_{02}(s)=\int_0^{\infty} e^{-st}\,\overline{H}_2(t)\,dF_1(t)=\frac{\lambda_1}{s+\lambda_1}[1-h_2(s+\lambda_1)],$$

where we denote the LS transforms of $G_i(t)$ and $H_i(t)$ by the corresponding small letters $g_i(s)$ and $h_i(s)$, respectively. In the second case, noting that the failure time of unit 1 has the memoryless property, we obtain

$$(4.2)\qquad g_{03}(s)=\int_0^{\infty} e^{-st}\,\overline{F}_1(t)\,dH_2(t)=h_2(s+\lambda_1).$$

In state S_1 we obtain each branch gain from the similar discussion of state S_0 as follows:

$$(4.3)\qquad g_{13}(s)=\int_0^{\infty} e^{-st}\,\overline{H}_1(t)\,dF_2(t)=\frac{\lambda_2}{s+\lambda_2}[1-h_1(s+\lambda_2)],$$

$$(4.4)\qquad g_{12}(s)=\int_0^{\infty} e^{-st}\,\overline{F}_2(t)\,dH_1(t)=h_1(s+\lambda_2).$$

In state S_2 we can consider the following two (exclusive and exhaustive) cases:

(i) The operative unit 2 fails before the repair completion of unit 1.

(ii) The repair of unit 1 is completed before the operative unit 2 fails.

In the first case the system goes to state S_4 (i.e., the system down). Its branch gain is

$$(4.5)\quad g_{24}(s) = \int_0^\infty e^{-st}\,\overline{G}_1(t)\,dF_2(t) = \frac{\lambda_2}{s+\lambda_2}\,[1 - g_1(s+\lambda_2)].$$

In the second case, noting that the failure time of unit 2 has the memoryless property, we obtain

$$(4.6)\quad g_{21}(s) = \int_0^\infty e^{-st}\,\overline{F}_1(t)\,dG_1(t) = g_1(s+\lambda_2).$$

From the similar discussion we obtain two branch gains in state S_3 :

$$(4.7)\quad g_{34}(s) = \int_0^\infty e^{-st}\,\overline{G}_2(t)\,dF_1(t) = \frac{\lambda_1}{s+\lambda_1}\,[1 - g_2(s+\lambda_1)],$$

$$(4.8)\quad g_{30}(s) = \int_0^\infty e^{-st}\,\overline{F}_1(t)\,dG_2(t) = g_2(s+\lambda_1).$$

We define $\varphi_0(s)$, the LS transform of the time distribution to first system down starting from state S_0 at $t = 0$. Using the results discussed in the preceding chapter, and defining that state S_0 is a source and state S_4 is a sink, we obtain immediately from Mason's gain formula

$$(4.9)\quad \varphi_0(s)=\frac{g_{02}g_{24}+g_{02}g_{21}g_{13}g_{34}+g_{03}g_{34}(1-g_{21}g_{12})}{1-g_{21}g_{12}-g_{03}g_{30}-g_{02}g_{21}g_{13}g_{30}+g_{21}g_{12}g_{03}g_{30}},$$

where we use the abbreviated notation g_{ij} instead of $g_{ij}(s)$ and each branch gain g_{ij} is given in (4.1)-(4.8). The mean time is given by

$$(4.10)\quad \hat{T}=-\left.\frac{d\varphi_0(s)}{ds}\right|_{s=0}.$$

4. 4. Special Cases

In this section we shall consider the three special cases of the results obtained in the preceding section. The first is the same system of identical units. The second is a two-unit standby redundant system without standby failure. The third is a two-unit paralleled redundant system.

The first special case is the same system of identical units. We assume that the failure time distribution of each operative unit is $F(t) = 1 - \exp(-\lambda t)$, the repair time distribution is $G(t)$, and the failure time distribution of each standby unit is $H(t)$.

The states of the system are defined as follows:

State S_0 ; a unit begins to be operative and the other remaining unit begins to be standby.

State S_1 ; a unit begins to be operative and the other remaining unit begins to get repaired.

State S_2 ; two units are under repair or failure, which denotes the system down.

The signal flow graph of the system is given in

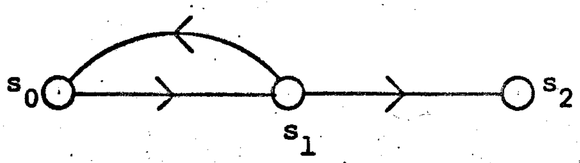

Fig. 4.2. Signal flow graph of the system of identical units.

Fig. 4.2. Each branch gain is given by

(4.11) $$g_{01}(s) = \int_0^\infty e^{-st}\,\overline{H}(t)\,dF(t) = \int_0^\infty e^{-st}\,\overline{F}(t)\,dH(t)$$

$$= \frac{\lambda}{s+\lambda}[1-h(s+\lambda)] + h(s+\lambda),$$

(4.12) $$g_{10}(s) = \int_0^\infty e^{-st}\,\overline{F}(t)\,dG(t) = g(s+\lambda),$$

(4.13) $$g_{12}(s) = \int_0^\infty e^{-st}\,\overline{G}(t)\,dF(t) = \frac{\lambda}{s+\lambda}[1-g(s+\lambda)].$$

Defining that state s_0 is a source and state s_2 is a sink, we obtain immediately from Mason's gain formula

(4.14) $$\varphi_0(s) = g_{01}(s)\,g_{12}(s) \,/\, [1 - g_{01}(s)\,g_{10}(s)].$$

The mean time is given by

$$(4.15)\qquad \hat{T} = \frac{1}{\lambda} + \frac{1 - h(\lambda)}{\lambda[1 - g(\lambda)]}.$$

These results can be also obtained by assuming that two units are identical in (4.9) and (4.10).

The second special case is a two-unit standby redundant system, where we assume that the failure of the standby unit never occurs in the standby interval. For the model, we should set $H_i(t) \equiv 0$ ($\bar{H}_i(t) \equiv 1$) in (4.1)-(4.8). In this case we obtain

(4.16)

$$\varphi_0(s) = \frac{\frac{\lambda_1}{s+\lambda_1}\cdot\frac{\lambda_2}{s+\lambda_2}\cdot\frac{\lambda_1}{s+\lambda_1}\, g_1(s+\lambda_2)\,[1-g_2(s+\lambda_1)] + \frac{\lambda_1}{s+\lambda_1}\cdot\frac{\lambda_2}{s+\lambda_2}[1-g_1(s+\lambda_2)]}{1-\frac{\lambda_1}{s+\lambda_1}\cdot\frac{\lambda_2}{s+\lambda_2}\, g_1(s+\lambda_2)\, g_2(s+\lambda_1)},$$

and

$$(4.17)\qquad \hat{T} = \frac{1}{\lambda_1} + \frac{1/\lambda_2 + g_1(\lambda_2)/\lambda_1}{1 - g_1(\lambda_2)g_2(\lambda_1)}.$$

These results have been given in (1.25) and (1.26) of Section 1.4. However, in the model we need not the memoryless property of the operative unit, i.e., the failure time distribution of the operative unit i (i = 1, 2) is an arbitrary one $F_i(t)$. In the most generalized model, we have given the LS transform by using the signal flow graph method in Section 3.4.

The third special case is a two-unit paralleled redundant system. In the results of Section 4.3, we assume that the failure time of the standby unit i (i = 1, 2) has also the same distribution $H_i(t) = F_i(t) = 1 - \exp(-\lambda_i t)$ as unit i is operative. For the model we obtain

$$(4.18) \quad \varphi_0(s) = \frac{\lambda_1\lambda_2\left[\frac{1-g_1(s+\lambda_2)}{s+\lambda_2} + \frac{1-g_2(s+\lambda_1)}{s+\lambda_1}\right]}{s+\lambda_1[1-g_1(s+\lambda_2)]+\lambda_2[1-g_2(s+\lambda_1)]},$$

and

$$(4.19) \quad \hat{T} = \frac{1+\frac{\lambda_1}{\lambda_2}[1-g_1(\lambda_2)]+\frac{\lambda_2}{\lambda_1}[1-g_2(\lambda_1)]}{\lambda_1[1-g_1(\lambda_2)]+\lambda_2[1-g_2(\lambda_1)]},$$

which have been given in Section 1.3.

In the second and third special cases we can also obtain the LS transforms for the same systems of identical units and their mean times by setting $H(t) \equiv 0$ for the second case and by setting $H(t) = 1 - \exp(-\lambda t)$ ($h(s) = \lambda/(s+\lambda)$) in (4.14) and (4.15).

4. 5. Analysis for the system with general failure

In the first analysis for the model we assume that the failure time distribution of the operative unit is exponential. In this section we assume that the failure time distribution of the operative unit i ($i = 1, 2$) is an arbitrary $F_i(t)$ ($t \geq 0$). We shall analyse the model by focussing on the regeneration point of the failure time distribution. First, we shall consider a simple system of identical units. The system of dissimilar units will be discussed in the following section.

For the simple system of identical unit we define the following three states of the system:

State S_0; one unit begins to be operative and another unit begins to be in standby.

State S_1; one unit begins to be operative and another unit begins to get repaired.

State S_2; two units are under repair or failure simultaneously. This state denotes the system down.

We note that these states denote the time instants (or epochs) of the model. The state transition diagram of the model is demonstrated in Fig. 4.3.

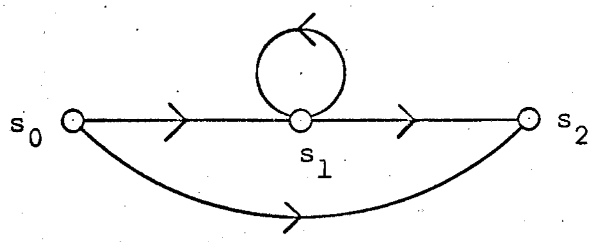

Fig. 4.3. The state transition diagram of the model of identical units.

We shall consider the state transitions from state S_0. In state S_0 two transitions can be considered; one is to state S_1, and another is to state S_2.

In the first place we shall consider the state transition from state S_0 to state S_1. We consider the time interval $(0, t)$. The probability that the operative unit fails first in the time interval $(t, t+dt)$ is $dF(t)$. In the time interval $(0, t)$, the probabilities that another unit is in standby up to time t are $\overline{H}(t)$, $H(t)*G(t)*\overline{H}(t)$, $H(t)*G(t)*H(t)*G(t)*\overline{H}(t)$, and so on, where $*$ denotes the convolution operation. We note that $\overline{H}(t)$ means that another unit never fails in $(0, t)$, $H(t)*G(t)*\overline{H}(t)$ means that another unit is in standby up to time t via the unit's failure and repair, $H(t)*G(t)*H(t)*G(t)*\overline{H}(t)$ means that another unit is in standby up to time t via two times of the unit's failure and repair, and so on. These events are mutually exclusive. Thus, the one step distribution (which may be improper) from state S_0 to state S_1 is

$$(4.20)\quad Q_{01}(t)=\int_0^t [\overline{H}(t)+H(t)*G(t)*\overline{H}(t)+H(t)*G(t)*H(t)*G(t)*\overline{H}(t) + \cdots\cdots]dF(t).$$

Introduce the notation

$$(4.21)\quad (1-A(t))^{(-1)}=\sum_{n=0}^{\infty}[A(t)]^{n*},$$

where

$$(4.22)\qquad [A(t)]^{n*}=\begin{cases}\overbrace{A(t)*A(t)*\cdots*A(t)}^{n} & (n\geq 1)\\ 1. & (n=0;\ \text{a Heaviside step function})\end{cases}$$

Using the notation (4.21), (4.20) can be rewritten

$$(4.23)\qquad Q_{01}(t)=\int_0^t [\overline{H}(t)*(1-H(t)*G(t))^{(-1)}]dF(t).$$

The LS transform of $Q_{01}(t)$ becomes

$$(4.24)\qquad q_{01}(s)=\int_0^\infty e^{-st}[\overline{H}(t)*(1-H(t)*G(t))^{(-1)}]dF(t).$$

In the second place we shall consider the state transition from state S_0 to state S_2. We also consider the time interval $(0,\ t)$. The probability that the operative unit fails first in the time interval $(t,\ t+dt)$ is $dF(t)$. In the time interval $(0,\ t)$, the probabilities that another unit is under repair up to time t are $H(t)*\overline{G}(t)$, $H(t)*G(t)*H(t)*\overline{G}(t)$, $H(t)*G(t)*H(t)*G(t)*H(t)*\overline{G}(t)$, and so on. Thus, the LS transform of the one step distribution from state S_0 to state S_2 is given by

$$(4.25)\qquad q_{02}(s)=\int_0^\infty e^{-st}[H(t)*\overline{G}(t)*(1-H(t)*G(t))^{(-1)}]dF(t).$$

We shall consider the state transitions from state S_1. In state S_1 two transitions can be considered; one is coming back to state S_1, and another is to state S_2.

In the first place we shall consider the state transition from state S_1 to state S_1 again. In this case we can consider that the operative unit fails in the time interval (t, $t + dt$) and another unit is in standby up to time t. The probabilities that another unit is in standby up to time t are $G(t) * \overline{H}(t)$, $G(t) * H(t) * G(t) * \overline{H}(t)$, and so on. Thus we have

$$(4.26) \qquad g_{11}(s) = \int_0^{\infty} e^{-st} \left[G(t) * \overline{H}(t) * (1 - H(t) * G(t))^{(-1)} \right] dF(t).$$

In the second place we shall consider the state transition from state S_1 to state S_2. In the similar way we shall consider that the operative unit fails in the time interval (t, $t + dt$) and another unit is under repair up to time t. Thus we have

$$(4.27) \qquad g_{12}(s) = \int_0^{\infty} e^{-st} \left[\overline{G}(t) * (1 - H(t) * G(t))^{(-1)} \right] dF(t).$$

We define $\varphi_i(s)$ (i = 0, 1), the LS transform of the time distribution to first system down starting from state S_i at $t = 0$. For $\varphi_0(s)$, we have by using $g_{ij}(s)$

$$(4.28) \qquad \varphi_0(s) = g_{02}(s) + g_{01}(s)\, \varphi_1(s).$$

For $\varphi_1(s)$, we have similarly

$$\varphi_1(s) = g_{12}(s) + g_{11}(s)\varphi_1(s). \tag{4.29}$$

Solving for $\varphi_0(s)$ in (4.28) and (4.29), we obtain

$$\varphi_0(s) = g_{02}(s) + g_{01}(s)\,g_{12}(s) / [1 - g_{11}(s)], \tag{4.30}$$

which is our desired result. The mean time is immediately given by

$$\hat{T} = -\left.\frac{d\varphi_0(s)}{ds}\right|_{s=0}. \tag{4.31}$$

4. 6. Dissimilar Unit Case

We have derived the LS transform for the simple model that the two units are identical. In this section we shall extend the model to one of dissimilar units.

We define the following four states of the model:

State S_0 ; unit 1 begins to be operative and unit 2 begins to be in standby.

State S_1 ; unit 1 begins to get repaired and unit 2 begins to be operative.

State S_2 ; unit 1 begins to be operative and unit 2 begins to get repaired.

State S_3 ; two unit are under repair or failure simultaneously. This state denotes the system down.

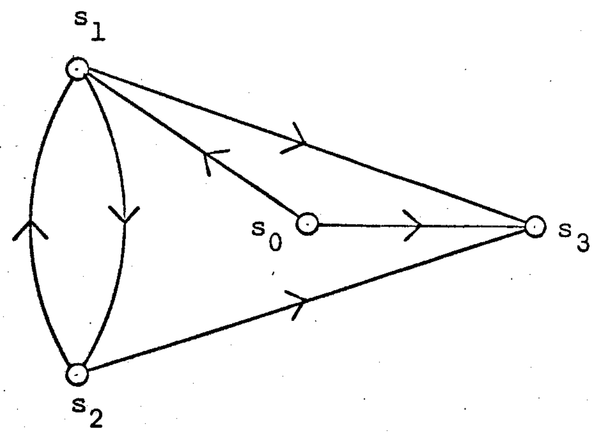

Fig. 4.4. The state transition diagram of the model of dissimilar units.

The state transition diagram is demonstrated in Fig. 4.4. Each LS transform $g_{ij}(s)$ of the one step distribution from state S_i to S_j is similarly given by

$$(4.32) \qquad g_{01}(s) = \int_0^\infty e^{-st}\left[\overline{H}_2(t) * (1 - H_2(t) * G_2(t))^{(-1)}\right] dF_1(t),$$

$$(4.33) \qquad g_{03}(s) = \int_0^\infty e^{-st}\left[H_2(t) * \overline{G}_2(t) * (1 - H_2(t) * G_2(t))^{(-1)}\right] dF_1(t),$$

$$(4.34) \qquad g_{12}(s) = \int_0^\infty e^{-st}\left[G_1(t) * \overline{H}_1(t) * (1 - H_1(t) * G_1(t))^{(-1)}\right] dF_2(t),$$

$$(4.35) \qquad g_{13}(s) = \int_0^\infty e^{-st}\left[\overline{G}_1(t) * (1 - H_1(t) * G_1(t))^{(-1)}\right] dF_2(t),$$

$$(4.36) \qquad g_{21}(s) = \int_0^\infty e^{-st}\left[G_2(t) * \overline{H}_2(t) * (1 - H_2(t) * G_2(t))^{(-1)}\right] dF_1(t),$$

(4.37) $$\gamma_{23}(s) = \int_0^{\infty} e^{-st} [\overline{G}_2(t) * (1 - H_2(t) * G_2(t))^{(-1)}] dF_1(t).$$

We also define $\varphi_i(s)$ (i = 0, 1, 2), the LS transform of the first time distribution to system down starting from state i. For each $\varphi_i(s)$, we have

(4.38) $$\varphi_0(s) = \gamma_{03}(s) + \gamma_{01}(s)\varphi_1(s),$$

(4.39) $$\varphi_1(s) = \gamma_{13}(s) + \gamma_{12}(s)\varphi_2(s),$$

(4.40) $$\varphi_2(s) = \gamma_{23}(s) + \gamma_{21}(s)\varphi_1(s).$$

Solving for $\varphi_0(s)$, we have

(4.41) $$\varphi_0(s) = \gamma_{03}(s) + \frac{\gamma_{01}(s)\gamma_{13}(s) + \gamma_{01}(s)\gamma_{12}(s)\gamma_{13}(s)}{1 - \gamma_{12}(s)\gamma_{21}(s)}$$

and the mean time can be derived from (4.31) by using $\varphi_0(s)$ in (4.41).

Two special cases can be easily obtained. The first case is a two-unit standby redundant system in no consideration of standby failure by setting $H_i(t) \equiv 0$ ($\overline{H}_i(t) \equiv 1$), the results of which have been given in Section 1.4. The second case is a two-unit paralleled redundant system by setting $F_i(t) \equiv H_i(t) \equiv 1 - \exp(-\lambda_i t)$, the results of which have been given in Section 1.3.

CHAPTER V

A Two-Unit Standby Redundant System WITH REPAIR AND PREVENTIVE MAINTENANCE

5. 1. Introduction

It is an important problem to operate a system in a specified long time without failure. We have known some policies to maintain a system. In particular, the following two policies are well-known:

(i) We make the system redundant.

(ii) We make the system preventively maintainable.

For the models using (i), a two-unit standby (or paralleled) redundant system is found in many fields and is well-known. The detailed discussion of such a system has been described in the preceding chapters. For the models using (ii), Barlow and Proschan [4, 5] have discussed as replacement problems. They have studied in detail a random replacement, an age replacement, a block replacement, and other replacement models.

In this chapter we shall consider a system which combines the above two policies. As a redundant model, we shall consider a two-unit standby redundant system with repair maintenance. As a preventive maintenance policy, we shall adopt a random one for an operative unit of the system. That is, an operative unit stops its operation after a time duration for the preventive maintenance. Combining the two policies just mentioned above, we call the system a two-unit standby redundant system with repair and preventive maintenance. Our concern for the system is the time to first system down.

First, we shall consider a system of two identical units. That is, we consider a system of two units in which the two units have the same statistical properties. Defining the states of the system, and focussing on the regeneration point of the failure or inspection time, we shall derive the Laplace-Stieltjes (LS) transform of the time distribution to first system down. The mean time will be also derived from it. We shall further show that the mean time derived here is greater than that of a two-unit standby redundant system with only repair maintenance under the suitable conditions.

Second, we shall consider a system of two dissimilar units. That is, we consider a system in which the statistical properties of the two units are different. In the same way we shall analyse the system by using the signal flow graph method.

5. 2. Model

Consider a system of two identical unit (or subsystems). The failure time distribution of each unit is an arbitrary $F(t)$ and the repair time distribution is also an arbitrary $G(t)$. We assume that after the repair completion a unit recovers its function perfectly. We also assume that the switchover times from the failure to the repair, from the repair completion to the standby state, and from the standby state to the operative state of each unit are all instantaneous. The behavior of the system obeys the usual two-unit standby redundant system (see Gnedenko et al. [35, p. 329] and Srinivasan [58].

Next we shall consider the preventive maintenance policy. When an operative unit goes to a specified time t and it is free from failure in that interval, the unit undergoes inspection as the preventive maintenance policy. We assume that the time distribution to the inspection is an arbitrary $A(t)$. The time distribution from the inspection to the inspection completion (or the preventive repair completion) is assumed to be an arbitrary $B(t)$. We assume that after the inspection completion a unit recovers its function perfectly. We also assume that $G(t) \leq B(t)$ for all t so as to make the preventive maintenance policy effective. We shall further consider a special situation: When an operative unit goes to the inspection time before the repair completion of the other failed unit (or the inspection completion of the other unit under inspection), we make no inspection for an operative unit since the inspection of the operative unit yields the system down. That is, the inspection of an operative unit is only made if the other unit is in standby. We assume that the switchover times occurring in the inspection are all instantaneous. We also assume that all random variables are mutually independent and nonnegative. We should naturally assume that the failure time distribution of an operative unit has IFR (see Barlow and Proschan [5, p. 12]) so as to make the preventive maintenance policy effective.

Our concern for the system is the LS transform of the time distribution to first system down. We shall derive the LS transform.

5. 3. Analysis

Consider the time instants of the failure or inspection of the units for the analysis of the system. We shall consider the following four states (which are the time instants of the system):

State S_0 : One unit begins to be operative and the other is in standby.

State S_1 : One unit begins to be operative instead of the other failed unit and the failed unit begins to get repaired.

State S_2 : One unit begins to be operative instead of the inspection of the other unit and the inspection of the other unit begins.

State S_3 : The two unit are under failure, inspection, or repair simultaneously, which state denotes the system down.

We shall consider the time distribution to first system down (i.e., state S_3) starting from state S_0 at $t = 0$. Then we shall consider each transition time distribution from one state to another.

In state S_0 , we can consider the following two (exclusive and exhaustive) cases:

(i) An operative unit fails before the inspection time comes.

(ii) The inspection time of an operative unit comes before an operative unit fails.

In case (i) the system goes to state S_1. Its distribution becomes

$$Q_{01}(t) = \int_0^t \overline{A}(t)\,dF(t), \tag{5.1}$$

where $\overline{A}(t) = 1 - A(t)$ denotes the survival probability function. In general, the upper bar of the distribution denotes the survival probability function throughout this chapter. Applying the LS transofrms for (5.1), we have

$$q_{01}(s) = \int_0^\infty e^{-st}\,dQ_{01}(t) = \int_0^\infty e^{-st}\,\overline{A}(t)\,dF(t). \tag{5.2}$$

In case (ii) the system goes to state S_2. The LS transform of the time distribution from state S_0 to state S_2 becomes

$$q_{02}(s) = \int_0^\infty e^{-st}\,\overline{F}(t)\,dA(t). \tag{5.3}$$

In state S_1, we consider the following three (exclusive and exhaustive) cases:

(i) After the repair completion of a failed unit, an operative unit fails.

(ii) After the repair completion of a failed unit, the inspection time comes.

(iii) An operative unit fails before the repair completion of a failed unit.

In case (i) we can further consider the following two (exclusive and exhaustive) cases: (A) After the repair completion of a failed unit, an operative unit fails and that the inspection time does not come in that interval. Then its distribution becomes $\int_0^t \bar{A}(t)G(t)\,dF(t)$. (B) The inspection time comes before the repair completion of a failed unit. In this case the inspection is not made as we have described in Section 5.2. Then the probability that the repair of a failed unit is completed up to time x after the inspection time comes is $\int_0^x A(y)\,dG(y)$. The time distribution that an operative unit fails after the repair completion (and that the inspection is not made) becomes $\int_0^t [\int_0^x A(y)\,dG(y)]\,dF(x)$. Thus we have the LS transform of the time distribution from state S_1 to state S_1 as follows:

$$(5.4)\qquad g_{11}(s)=\int_0^\infty e^{-st}\bar{A}(t)G(t)\,dF(t)+\int_0^\infty e^{-st}[\int_0^t A(y)\,dG(y)]\,dF(t).$$

In case (ii), after the repair completion of a failed unit, the inspection time comes and it is free from failure of an operative unit in that interval. Then the system goes to state S_2. Its LS transform becomes

$$(5.5) \qquad \gamma_{12}(s) = \int_0^\infty e^{-st}\,\overline{F}(t)\,G(t)\,dA(t).$$

In case (iii) the system goes to state S_3. Its LS transform becomes

$$(5.6) \qquad \gamma_{13}(s) = \int_0^\infty e^{-st}\,\overline{G}(t)\,dF(t).$$

In state S_2, we can consider the following three (exclusive and exhaustive) cases:

(i) After the inspection is completed, the inspection time of an operative unit comes.

(ii) After the inspection is completed, an operative unit fails.

(iii) An operative unit fails before the inspection is completed.

In case (i), after the inspection is completed, the inspection time of an operative unit comes and it is free from failure of an operative unit in that interval. Then the system goes to state S_2. Its LS transform becomes

$$(5.7) \qquad \gamma_{22}(s) = \int_0^\infty e^{-st}\,\overline{F}(t)\,B(t)\,dA(t).$$

In case (ii) we can further consider the following two (exclusive and exhaustive) cases: (A) An operative unit fails after the inspection completion and that the inspection time does not come in that interval. (B) The inspection time of an operative unit comes before the inspection completion. In this case the inspection is not made so as to avoid the system down. Then the inspection is completed before an operative unit fails. In both cases (A) and (B), the system goes to state S_1. In the similar way of deriving (5.4), we have

$$(5.8) \qquad g_{21}(s) = \int_0^\infty e^{-st}\overline{A}(t)\,B(t)\,dF(t) + \int_0^\infty e^{-st}\left[\int_0^t A(y)\,dB(y)\right]dF(t).$$

In case (iii), the system goes to state S_3 (i.e., the system down). Its LS transform becomes

$$(5.9) \qquad g_{23}(s) = \int_0^\infty e^{-st}\,\overline{B}(t)\,dF(t).$$

Thus we have all branch gains of the signal flow graph of the system. We show the signal flow graph of the system in Fig. 5.1, where each branch gain is given by (5.2)-(5.9). Defining that state S_0 is a source and state S_3 is a sink in Fig. 5.1, and applying Mason's gain formula [13], we have

$$(5.10) \quad \varphi_0(s) = \frac{g_{01}(s)\,g_{13}(s)[1-g_{22}(s)] + g_{01}(s)\,g_{12}(s)\,g_{23}(s) + g_{02}(s)\,g_{23}(s)[1-g_{11}(s)] + g_{02}(s)\,g_{21}(s)\,g_{13}(s)}{1 - g_{11}(s) - g_{22}(s) + g_{11}(s)\,g_{22}(s) - g_{12}(s)\,g_{21}(s)},$$

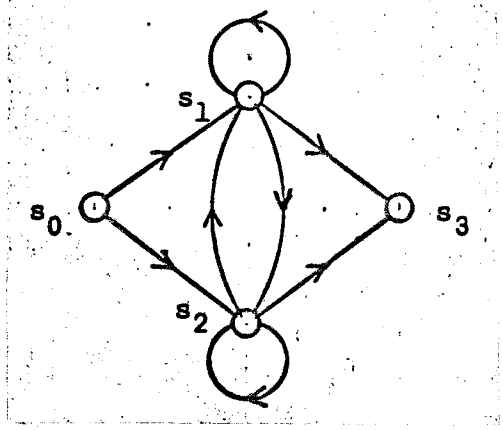

Fig. 5.1. Signal flow graph of the system of identical units.

which is the LS transform of the time distribution to first system down starting from state S_0 at $t = 0$.

To prove that the above distribution is a proper one [28, p. 129], we should prove $\varphi_0(0) = 1$. Then we should verify

(5.11) $\qquad g_{01}(0) + g_{02}(0) = 1,$

(5.12) $\qquad g_{11}(0) + g_{12}(0) + g_{13}(0) = 1,$

(5.13) $\qquad g_{21}(0) + g_{22}(0) + g_{23}(0) = 1.$

As we have described in deriving $g_{ij}(s)$, we have considered all the possibilities in state S_i (i = 0, 1, 2). Thus we have verified that (5.11), (5.12), and

(5.13) hold. We can also verify analytically that (5.11), (5.12), and (5.13) hold, but we omit the proof.

5. 4. Mean Time and Discussions

In this section we shall derive the mean time to first system down. We shall further discuss some properties concerning the mean time. To simplify the notation, we introduce the notations

$$(5.14) \qquad \gamma_{ij} \equiv \gamma_{ij}(0), \qquad (i, j = 0, 1, 2, 3)$$

$$(5.15) \qquad \mu_0 \equiv -\frac{d\gamma_{01}(s)}{ds}\bigg|_{s=0} - \frac{d\gamma_{02}(s)}{ds}\bigg|_{s=0},$$

$$(5.16) \qquad \mu_i \equiv -\sum_{j=1}^{3} \frac{d\gamma_{ij}(s)}{ds}. \qquad (i = 1, 2)$$

Using the above notations (5.14), (5.15), and (5.16), we have the mean time to first system down

$$(5.17) \qquad \hat{T} = -\frac{d\varphi_0(s)}{ds}\bigg|_{s=0}$$

$$= \mu_0 + \frac{(\gamma_{01}\gamma_{13} + \gamma_{21})\mu_1 + (1 - \gamma_{11} - \gamma_{01}\gamma_{13})\mu_2}{(1 - \gamma_{11})\gamma_{13} + \gamma_{21}\gamma_{13}}.$$

We have discussed a random preventive maintenance policy. We further consider an age preventive maintenance policy. In practical situations we should adopt an age preventive maintenance policy since it is suitable for the actual policy. Then we assume that

$$A(t) = \begin{cases} 0 & \text{for } t < t_0 \\ 1 & \text{for } t \geq t_0. \end{cases} \tag{5.18}$$

In this case we introduce the following notations

$$\theta_1 \equiv \int_0^\infty \bar{G}(t)dF(t), \qquad \theta_2 \equiv \int_0^\infty \bar{B}(t)dF(t), \tag{5.19}$$

$$\beta_1 \equiv F(t_0), \qquad \beta_2 \equiv \bar{F}(t_0), \tag{5.20}$$

$$\gamma \equiv \int_0^{t_0} \bar{F}(t)dt, \tag{5.21}$$

$$1/\lambda \equiv \int_0^\infty F(t)dt = \int_0^t t dF(t), \tag{5.22}$$

$$G \equiv G(t_0), \qquad B \equiv B(t_0). \tag{5.23}$$

Using the above notations (5.19)-(5.23), we have for an age preventive maintenance policy

$$\hat{T} = \frac{1+\theta_1}{\lambda\theta_1} + \frac{(\theta_1+G)\{\theta_1\gamma-(\beta_1\theta_1+\beta_2\theta_2)/\lambda\}}{\theta_1\{\theta_1+\beta_2(G\theta_2-B\theta_1)\}}, \tag{5.24}$$

where the first term of the above equation denotes the mean time without the preventive maintenance and the second term denotes the effect of the preventive maintenance.

The results (5.29) is equal to the first term of the right-hand side of equation (5.24) as is shown above. Thus the second term of equation (5.24) denotes the effect of the preventive maintenance policy.

We shall finally discuss the following theorem that the preventive maintenance policy is effective in the sense of the mean time.

Theorem 5.1. The mean time (5.24) for the system with repair and preventive maintenance is greater than that of (5.29) for the system with only repair maintenance on the assumptions that the failure rate $r(t)$ of the failure time distribution is strictly increasing and there exists a t_0^* such that

$$r(t_0^*) = \lambda\theta_1/(\theta_1-\theta_2), \tag{5.30}$$

and that we adopt a suitable inspection interval .

Proof. To prove the theorem, we should verify that the second term of the right-hand side of equation (5.24) is positive on the above assumptions. We should only consider the second term of the right-hand side of equation (5.24). We shall only show that the denominator and the numerator of the second term are both positive on the above assumptions. It is evident from (5.19) and (5.23) that θ_1 and $\theta_1 + G$ are both positive. The brackets of the denominator become

$$\begin{aligned} &\theta_1 + \beta_2(G\theta_2 - B\theta_1) \\ &= (1-\bar{F}(t_0)B(t_0))\theta_1 + \bar{F}(t_0)G(t_0)\theta_1 > 0, \end{aligned} \tag{5.31}$$

As a special case of the system discussed in this chapter, we shall consider a two-unit standby redundant system with only repair maintenance. In this case we may only consider the states S_0, S_1, and S_3 which are defined in Section 5.3. We need not to consider state S_2 since the inspection is not made. Each LS transform of the transidion time distribution from one state to another is given by

(5.25) $$\xi_{01}(s) = \int_0^\infty e^{-st}\, dF(t),$$

(5.26) $$\xi_{11}(s) = \int_0^\infty e^{-st}\, G(t)\, dF(t),$$

(5.27) $$\xi_{13}(s) = \int_0^\infty e^{-st}\, \overline{G}(t)\, dF(t).$$

These results can also be obtained by setting $A(t) \equiv 0$ ($\overline{A}(t) \equiv 1$) for all t in (5.2)-(5.9). The LS transform of the time distribution to first system down is given by

(5.28) $$\varphi_0(s) = \frac{\xi_{01}(s)\, \xi_{13}(s)}{1 - \xi_{11}(s)},$$

where each $\xi_{ij}(s)$ is defined in (5.25)-(5.27). The mean time to first system down is given by

(5.29) $$\hat{T} = \frac{1+\theta_1}{\lambda\theta_1}.$$

from $\overline{F}(t_0) < 1$ and $B(t_0) < 1$. Define $p(t_0)$ by the brackets of the numerator, which is a function of t_0.

$$(5.32)\quad p(t_0) = \theta_1 \gamma - (\beta_1\theta_1 + \beta_2\theta_2)/\lambda$$
$$= \theta_1 \int_0^{t_0} \overline{F}(t)dt - \{(\theta_1 - \overline{F}(t_0))\theta_1 + \overline{F}(t_0)\theta_2\}/\lambda$$
$$= (\theta_1 - \theta_2)\overline{F}(t_0)/\lambda + \theta_1 \int_{t_0}^{\infty} \overline{F}(t)dt.$$

For $p(t_0)$, we have $p(0) = -\theta_2/\lambda < 0$, $p(\infty) = 0$. Differentiating $p(t_0)$ with respect to t_0, we have

$$(5.33)\quad \frac{dp(t_0)}{dt} = -(\theta_1 - \theta_2)f(t_0)/\lambda + \theta_1\overline{F}(t_0)$$
$$= \overline{F}(t_0)\{\theta_1 - (\theta_1 - \theta_2)r(t_0)/\lambda\},$$

where $f(t_0) = dF(t_0)/dt_0$ and $r(t_0) = f(t_0)/\overline{F}(t_0)$. By using the assumptions that $r(t_0)$ is an increasing function of t_0 (i.e., $r(t_0)\uparrow$ as $t_0\uparrow$, we can show that there exists a t_0^* such that $dp(t_0)/dt_0 = 0$, that is

$$(5.34)\quad r(t_0^*) = \frac{\lambda\theta_1}{\theta_1 - \theta_2},$$

where $\theta_1 - \theta_2 > 0$ if $B(t) \geq G(t)$ and $G(t) \not\equiv B(t)$ for all t. Using also $p(0) < 0$, $p(\infty) = 0$, and $p(t_0)$ is a unimodal function of $t_0 > 0$, there exists a $\overline{t}_0$ such that $p(\overline{t}_0) = 0$. Thus, if we choose a $t_0 > \overline{t}_0$, we have $p(t_0) > 0$. That is, the second term of the right-hand side of equation (5.24) is positive if we choose a $t_0\ (> \overline{t}_0)$, which proves the theorem.

5. 5. Dissimilar Case

In this section we shall further consider a two-unit standby redundant system with repair and preventive maintenance in which the two units are different in their statistical properties. We shall simply describe the necessary definitions of the system.

The two units can be labeled by the integers i = 1, 2. The failure time distribution of unit i (i = 1, 2) is an arbitrary $F_i(t)$ and the repair time distribution of unit i is also an arbitrary $G_i(t)$. The time distribution from the beginning of the operation to the inspection of the operative unit i is also an arbitrary $A_i(t)$. The time distribution from the inspection to the inspection completion (or the preventive repair completion) of unit i under inspection is an arbitrary $B_i(t)$. We also assume that all random variables are mutually independent and nonnegative. The same assumptions of the system are imposed as described in Section 5. 2. For example, we should assume that $F_i(t)$ (i = 1, 2) has IFR, $B_i(t) \geq G_i(t)$ (i = 1, 2), and so on.

Consider the time instants of the failure or inspection of the units for the analysis of the system. We shall consider the following six states (which are the time instants of the system):

State S_0: Unit 1 begins to be operative and unit 2 begins to be in standby.

State S_1: Unit 2 begins to be operative instead of the failed unit 1 and the failed unit 1 begins to get repaired.

State S_2: Unit 1 begins to be operative instead of the failed unit 2 and the failed unit 2 begins to get repaired.

State S_3: Unit 2 begins to be operative instead of the other unit 1 and the inspection of unit 1 begins.

State S_4: Unit 1 begins to be operative instead of the other unit 2 and the inspection of unit 2 begins.

State S_5: The two units are under failure, inspection, or repair simultaneously, which state denotes the system down.

We shall consider each LS transform of the transition time distribution from one state to another. Each LS transform is derived in the similar way of the system of identical units. Then we have

(5.35) $$\gamma_{01}(s)=\int_0^\infty e^{-st}\,\overline{A}_1(t)\,dF_1(t),$$

(5.36) $$\gamma_{03}(s)=\int_0^\infty e^{-st}\,\overline{F}_1(t)\,dA_1(t),$$

(5.37) $$\gamma_{12}(s)=\int_0^\infty e^{-st}\,\overline{A}_2(t)\,G_1(t)\,dF_2(t)+\int_0^\infty e^{-st}\left[\int_0^t A_2(y)\,dG_1(y)\right]dF_2(t),$$

(5.38) $$\gamma_{14}(s)=\int_0^\infty e^{-st}\,\overline{F}_2(t)\,G_1(t)\,dA_2(t),$$

(5.39) $g_{15}(s)=\int_0^\infty e^{-st}\,\overline{G}_1(t)\,dF_2(t),$

(5.40) $g_{21}(s)=\int_0^\infty e^{-st}\,\overline{A}_1(t)\,G_2(t)\,dF_1(t)+\int_0^\infty e^{-st}\left[\int_0^t A_1(y)\,dG_2(y)\right]dF_1(t),$

(5.41) $g_{23}(s)=\int_0^\infty e^{-st}\,\overline{F}_1(t)\,G_1(t)\,dA_2(t),$

(5.42) $g_{25}(s)=\int_0^\infty e^{-st}\,\overline{G}_2(t)\,dF_1(t),$

(5.43) $g_{34}(s)=\int_0^\infty e^{-st}\,\overline{F}_2(t)\,B_1(t)\,dA_2(t),$

(5.44) $g_{32}(s)=\int_0^\infty e^{-st}\,\overline{A}_2(t)B_1(t)\,dF_2(t)+\int_0^\infty e^{-st}\left[\int_0^t A_2(y)\,dB_1(y)\right]dF_2(t),$

(5.45) $g_{35}(s)=\int_0^\infty e^{-st}\,\overline{B}_2(t)\,dF_1(t),$

(5.46) $g_{43}(s)=\int_0^\infty e^{-st}\,\overline{F}_1(t)\,B_2(t)\,dA_1(t),$

(5.47) $g_{41}(s)=\int_0^\infty e^{-st}\,\overline{A}_1(t)\,B_2(t)\,dF_1(t)+\int_0^\infty e^{-st}\left[\int_0^t A_1(y)\,dB_2(y)\right]dF_1(t),$

(5.48) $g_{45}(s)=\int_0^t e^{-st}\,\overline{B}_1(t)\,dF_2(t),$

The signal flow graph of the system is demonstrated in Fig. 5.2. Each brach gain is given in (5.35)-(5.48).

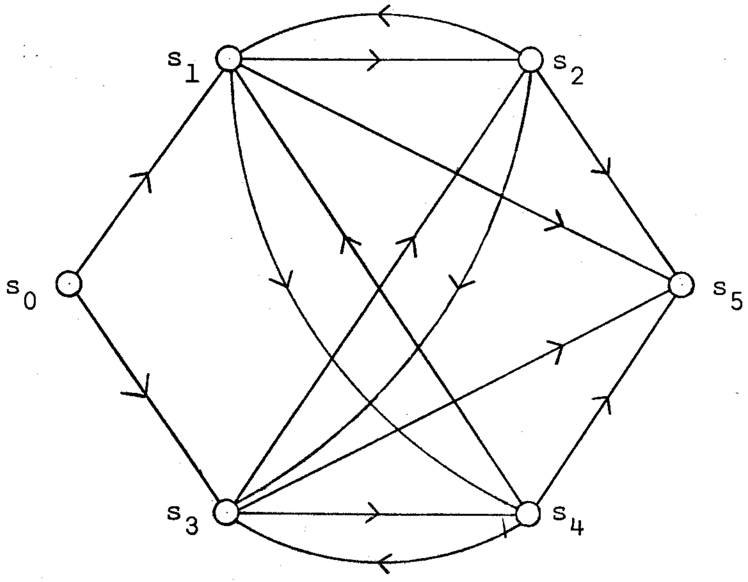

Fig. 5.2. Signal flow graph of the system of dissimilar units.

Thus, defining that state s_0 is a source and state s_5 is a sink, and applying Mason's gain formula, we have immediately

(5.49) $$\varphi_0(s) = N/D,$$

where

(5.50) $$D = 1 - g_{12}(s)g_{21}(s) - g_{14}(s)g_{41}(s) - g_{34}(s)g_{43}(s) - g_{23}(s)g_{32}(s)$$
$$+ g_{12}(s)g_{21}(s)g_{34}(s)g_{43}(s) + g_{14}(s)g_{41}(s)g_{23}(s)g_{32}(s),$$

$$
\begin{aligned}
(5.51)\quad N = {} & g_{01}(s)\, g_{12}(s)\, g_{15}(s)\, [1 - g_{34}(s)\, g_{43}(s)] + g_{01}(s)\, g_{12}(s)\, g_{23}(s)\, g_{34}(s)\, g_{45}(s) \\
& + g_{01}(s)\, g_{15}(s)\, [1 - g_{34}(s)\, g_{43}(s) - g_{23}(s)\, g_{32}(s)] + g_{01}(s)\, g_{12}(s)\, g_{23}(s)\, g_{35}(s) \\
& + g_{01}(s)\, g_{14}(s)\, g_{45}(s)\, [1 - g_{23}(s)\, g_{32}(s)] \\
& + g_{03}(s)\, g_{34}(s)\, g_{45}(s)\, [1 - g_{12}(s)\, g_{21}(s)] + g_{03}(s)\, g_{34}(s)\, g_{41}(s)\, g_{12}(s)\, g_{25}(s) \\
& + g_{03}(s)\, g_{35}(s)\, [1 - g_{12}(s)\, g_{21}(s) - g_{14}(s)\, g_{41}(s)] + g_{03}(s)\, g_{34}(s)\, g_{41}(s)\, g_{15}(s) \\
& + g_{03}(s)\, g_{32}(s)\, g_{35}(s)\, [1 - g_{14}(s)\, g_{41}(s)].
\end{aligned}
$$

We have obtained $\varphi_0(s)$, the LS transform of the time distribution to first system down. The mean time can be easily obtained from it. The similar result of Theorem 5.1 will hold for the system, but we omit the form of the theorem here.

5. 6. Conclusion

We have considered a two-unit standby redundant system with repair and preventive maintenance. For the system we have obtained the LS transform of the time distribution to first system down and its mean time. We have further shown that the preventive maintenance policy is effective in the sense of the mean time on the suitable assumptions. For the failure, repair, and inspection time distributions, we have assumed arbitrary distributions. Thus, our results obtained in this chapter are available by assuming suitable distributions.

In a recent paper, Mine and Asakura[†] have discussed a multiple unit standby redundant system with repair and preventive maintenance. They have derived the LS transform of the time distribution to the first emptiness and its mean time on the assumptions that the failure and inspection time distributions are exponential. In this chapter, we have derived the LS transform on the assumptions that all distributions are arbitrary, where we have considered a two-unit standby redundant system.

In many fields we use a two-unit standby redundant system. In this situations, if the failure time distribution has IFR, we should adopt the preventive maintenance policy. Then Theorem 5.1 states that the preventive maintenance policy is effective on the suitable assumptions. In the actual situations, these assumtions may be satisfied. In Theorem 5.1, we have adopted an age preventive maintenance policy. However, we believe that a random preventive maintenance policy (which includes an age one) is effective on the suitable assumption of the random inspection distribution.

---

† Mine, H. and Asakura, T., "The effect of an age replacement to a standby redundant system," J. Appl. Prob., vol. 6 (1969), no. 4, to appear.

CHAPTER VI

MARKOVIAN DECISION PROCESSES WITH DISCOUNTING

6. 1. Introduction

Consider a system whose state space has finitely many states as we have described in the Introduction. Let a state space S be a set of states labeled by the integers $i = 1, 2, \ldots, N$. That is, $S = \{1, 2, \ldots, N\}$. For each $i \in S$, we have a set K_i of finite actions (or alternatives) labeled by the integers $k = 1, 2, \ldots, K_i$. The policy space is denoted by the Cartesian product of each action set, i.e., $K = K_1 \times K_2 \times \ldots \times K_N$. Then we consider a sequential decision problem, i.e., we observe periodically one of states at time $n = 0, 1, 2, \ldots,$ and have to make an action at each time.

When the system is in state $i \in S$ and that we make an action $k \in K_i$, the following two things happen:

(i) We have the return r_i^k.

(ii) The system obeys the probability law p_{ij}^k $(j \in S)$ at the next time, where p_{ij}^k is the transition probability that the system is in state j at the next time given that the system is in state i at that time and an action k is made.

Here we assume that the return r_i^k is bounded for all $i \in S$, $k \in K_i$. It is clear from the finiteness of state space that

$$(6.1) \qquad \sum_{j \in S} p_{ij}^k = 1, \quad p_{ij}^k \geq 0. \qquad (i, j \in S,\ k \in K_i)$$

In this chapter, we consider a discounted process. Let β $(0 \le \beta < 1)$ be a discount factor. That is, unit return becomes β^n after n times (e.g., n days). The discount factor is considered as the reciprocal of 1 plus the interest rate. The introduction of the discount factor is mathematically to avoid the divergence of the total expected return.

We also give an initial distribution

$$a = (a_1, a_2, \dots, a_N), \tag{6.2}$$

where

$$\sum_{i \in S} a_i = 1, \qquad a_i \ge 0. \qquad (i \in S) \tag{6.3}$$

Then the system is a nonstationary Markov chain with returns. Our problem is to find strategies which maximizes the discounted total expected return over a finite or an infinite time horizon, where a strategy is a sequence of decisions in each time and each state.

In this chapter we shall focus our attention on a discounted decision process over an infinite time horizon. We consider a maximization problem (if we consider a minization problem, we may change the sign of the returns).

6. 2. Policy Iteration Algorithm

In this section we shall give the so-called Policy Iteration Algorithm for a discounted Markovian decision process. The policy iteration algorithm has been given by Howard [40]. So it is sometimes called Howard's policy iteration algorithm. The policy iteration algorithm has the close relation to linear programming discussed afterwards.

Since we consider a sequential decision process, we have the complete information of all states and times. Let F be a set of functions from the state space S to the policy space K. Since S and its associated K are both finite sets, F is a finite set. Let f or g be a function in F. Then a strategy π is defined by a sequence $\{f_n, n = 1, 2, \ldots\}$. So, we may write a strategy $\pi = (f_1, f_2, \ldots, f_n, \ldots)$, where f_n is the decision vector for each state at time n, i.e., $f_n(i)$, the ith element of f_n, is an action of state $i \in S$ at time n.

A strategy $(g, f_1, f_2, \ldots)$ is denoted by (g, π), where $g \in F$ and $\pi = (f_1, f_2, \ldots)$. A strategy $(f, f, \ldots, f, \ldots)$ is denoted by f^∞, where $f \in F$, and we call it a stationary strategy. That is, a stationary strategy f^∞ is one which is independent of time n. A strategy $(\overbrace{g, g, \ldots, g}^{n}, f_1, f_2, \ldots)$ is denoted by (g^n, π), where $g \in F$ and $\pi = (f_1, f_2, \ldots)$.

For any strategy π, we have a nonstationary Markov chain. Then we write n step transition probability matrix as

$$(6.4) \qquad P_n(\pi) = P(f_1)P(f_2)\cdots P(f_n). \qquad (n = 1, 2, \ldots)$$

where $P(f_n)$ is the $N\times N$ transition matrix whose i-j th element is p_{ij}^k, $k = f_n(i) \in K_i$. For $n = 0$, we define $P_0(\pi) = I$ (the $N\times N$ identity matrix). For any $f\in F$, we may write the $N\times 1$ return vector $r(f)$ whose ith element is r_i^k, $k = f(i) \in K_i$. Under the notation defined above, we have the $N\times 1$ discounted total expected return vector starting in each state $i\in S$:

$$(6.5) \qquad V_\beta(\pi) = \sum_{n=0}^{\infty} \beta^n P_n(\pi)\, r(f_{n+1}).$$

To show the finiteness of the above vector, setting $r^U = \max_{i,k} r_i^k$ and $r^L = \min_{i,k} r_i^k$, we have

$$(6.6) \qquad \frac{r^L}{1-\beta}\mathbb{1} \le V_\beta(\pi) \le \frac{r^U}{1-\beta}\mathbb{1},$$

where $\mathbb{1}$ is the $N\times 1$ vector with all elements unity.

While we have

$$(6.7) \qquad \begin{aligned} V_\beta(\pi) &= \sum_{n=0}^{\infty} \beta^n P_n(\pi)\, r(f_{n+1}) \\ &= r(f_1) + \sum_{n=1}^{\infty} \beta^n P_n(\pi)\, r(f_{n+1}) \\ &= r(f_1) + \beta P(f_1) \sum_{n=0}^{\infty} \beta^n P_n(T\pi)\, r(f_{n+2}) \\ &= r(f_1) + \beta P(f_1)\, V_\beta(T\pi), \end{aligned}$$

where $T\pi = (f_2, f_3, \ldots)$ is a strategy delayed one step for each time. Then, for any $N\times 1$ vector w we define a function $L(f)$ which maps w into $L(f)w = r(f) + \beta P(f)w$. We may consider that a function $L(f)$

is one step return under the preceding return w when we use the decision $f \in F$. Thus, $V_\beta(f, \pi) = L(f)\, V_\beta(\pi)$.

We define the vector inequality as follows. For any two column vectors w_1, w_2, we write $w_1 \geq w_2$ if every element of w_1 is not less than the corresponding element of w_2, and $w_1 > w_2$ if $w_1 \geq w_2$ and $w_1 \neq w_2$. This notation of vector inequality is used throughout this thesis.

Definition 6. 1. A strategy π^* is called β-optimal if $V_\beta(\pi^*) \geq V_\beta(\pi)$ for all π, where β $(0 \leq \beta < 1)$ is fixed.

This definition means that an optimal strategy is attained simultaneously for each initial state. But this fact is nontrivial, and will be shown below. If an optimal strategy π^* is attained simultaneously for each initial state, we have $a\, V_\beta(\pi^*) \geq a\, V_\beta(\pi)$ for any π and any initial distribution a, i.e., an optimal strategy is independent of the initial distribution a. Conversely, if an optimal strategy which maximizes $a\, V_\beta(\pi)$ is independent of the initial distribution a, we have $V_\beta(\pi^*) \geq V_\beta(\pi)$.

Lemma 6. 2. $L(f)$ is monotone, i.e., $w_1 \geq w_2$ implies $L(f)\, w_1 \geq L(f)\, w_2$.

Proof. $L(f)\, w_1 - L(f)\, w_2 = \beta\, P(f)(w_1 - w_2) \geq 0$ if $w_1 \geq w_2$, which completes the proof.

Under the above preparation, we have the following three theorems which have been given by Blackwell [9].

Theorem 6. 3. $V_\beta(\pi^*) \geq V_\beta(g, \pi^*)$ for all $g \in F$ implies that π^* is β-optimal.

Theorem 6. 4. $V_\beta(f, \pi) > V_\beta(\pi)$ implies $V_\beta(f^\infty) > V_\beta(\pi)$.

Now we have our main theorem.

Theorem 6. 5. Take any $f \in F$. For each $i \in S$, denote $G(i, f)$ the set of all $k \in K_i$ for which

$$r_i^k + \beta \sum_{j \in S} p_{ij}^k v_j > v_i ,$$

where v_i is the ith element of $V_\beta(f^\infty)$. If $G(i, f)$ is empty for all $i \in S$, then f^∞ is β-optimal. For any $g \in F$ such that

(a) $g(i) \in G(i, f)$ for some i and

(b) $g(i) = f(i)$ whenever $g(i) \notin G(i, f)$,

we have $V_\beta(g^\infty) > V_\beta(f^\infty)$.

As a direct consequence of the above theorems, we have

Corollary 6. 6. There is a β-optimal strategy which is stationary.

These theorems describe a method for finding an optimal stationary strategy, which is called (Howard's) Policy Iteration Algorithm. The algorithm has two parts as follows:

Value Determination Operation

Take any $f \in F$. Solve

$$v_i = r_i^k + \beta \sum_{j \in S} p_{ij}^k v_j$$

for $v_i \, (i \in S)$, where $k = f(i)$ correspods to the chosen strategy f^∞.

Policy Improvement Routine

Using the values $v_i \, (i \in S)$, find the element of $G(i, f)$ for each $i \in S$ such that

$$r_i^k + \beta \sum_{j \in S} p_{ij}^k v_j > v_i$$

for all $k \in K_i$. If $G(i, f)$ is empty for all $i \in S$, f^∞ is β-optimal and $V_\beta(f^\infty) = [\, v_i \,]$ is the discounted total expected return. If al least $g(i) \in G(i, f)$ for some i, make an improved strategy g^∞ such that $g(i) \in G(i, f)$ for some i and $g(i) = f(i)$ for $G(i, f)$ empty and return to Value Determination Operation.

As an initial strategy f^∞, we may take, for example, $\max_{k \in K_i} r_i^k$ for each $i \in S$.

This policy iteration algorithm is simple and elegant. The numerical examples using this algorithm will be presented in Section 6. 6. The convergence to an optimal strategy is fairly rapid. In Section 6. 4 we shall discuss the setup of this algorithm from the viewpoint of linear programming.

6. 3. Linear Programming Algorithm

In this section we shall consider the linear programming formulation of the discounted Markovian decision processes. The relation between the policy iteration algorithm and a linear programming algorithm will be described in the next section. In this section we formulate a linear programming algorithm and discuss some interesting properties of this problem.

It is more convenient for the latter discussion to extend the range of decisions to include randomized (or mixed) strategies. Thus for any n, we define $\lambda_i^k(n)$, the joint probability of being in state $i \in S$ and making decision $k \in K_i$. Our problem then is to find an optimal strategy which maximizes the discounted total expected return. Here we consider the maximization problem under the initial distribution (6.2) because an optimal strategy is attained simultaneously for each state.

Since $\lambda_i^k(n)$ obeys the probability law p_{ij}, we may write

$$(6.9) \quad \sum_{k \in K_j} \lambda_j^k(n) = \lambda_j(0) = a_j \qquad (n = 0)$$

$$= \sum_{i \in S} \sum_{k \in K_i} p_{ij}^k \lambda_i^k(n-1), \quad (n = 1, 2, \ldots; \ j \in S)$$

where a_j is the probability that the system is in state j at time 0 defined in (6.2).

Lemma 6. 7. Any nonnegative solution $\lambda_i^k(n)$ of (6.9) is a probability distribution and the corresponding discounted total expected return is bounded.

Proof. Since $\sum_{j \in S} p_{ij}^k = 1$, summing (6.9) over all $j \in S$ implies

$$\text{(6.10)} \quad \sum_{j \in S} \sum_{k \in K_j} \lambda_j^k(n) = \sum_{j \in S} a_j \qquad (n = 0)$$

$$= \sum_{i \in S} \sum_{k \in K_i} \lambda_i^k(n-1). \qquad (n = 1, 2, \ldots)$$

The assumption that a_j is a probability distribution implies recursively

$$\text{(6.11)} \quad \sum_{i \in S} \sum_{k \in K_i} \lambda_i^k(n) = 1. \qquad (n = 1, 2, \ldots)$$

This plus the fact that $\lambda_i^k(n)$ is assumed to be non-negative, which proves the first part of Lemma. The second part has been proved in (6.6). These complete the proof.

As a result of this lemma, we have the following objective function:

(6.12) $\quad \text{Max} \sum_{n=0}^{\infty} \beta^n \sum_{i \in S} \sum_{k \in K_i} r_i^k \lambda_i^k(n)$

under the constraints (6.9) and $\lambda_i^k(n) \geq 0$ for all $n = 0, 1, \ldots$; $i \in S$, and $k \in K_i$. As it stands, this problem, say problem (P_0), is similar to a standard linear programming problem; however, it contains an infinite number of constraints and variables and thus the classical theory of linear programming cannot be used to analyze it in this form. Using the fact that the sequences $\lambda_j^k(n)$ are bounded, and $0 \leq \beta < 1$, we define a set of new variables x_j^k by

(6.13) $$x_j^k = \sum_{n=0}^{\infty} \beta^n \lambda_j^k(n). \qquad (j \in S,\ k \in K_j)$$

By definition, the new variables x_j^k can be viewed as the z transforms of the sequences $\lambda_j^k(n)$ evaluated at $z = \beta$. Using these variables x_j^k, we have the following standard linear programming problem, say problem (P_T):

(6.14) $\quad \text{Max} \sum_{j \in S} \sum_{k \in K_j} r_j^k x_j^k$

subject to

(6.15) $$\sum_{k \in K_j} x_j^k + \beta \sum_{i \in S} \sum_{k \in K_i} p_{ij}^k x_i^k = a_j \qquad (j \in S)$$

(6.16) $$x_j^k \geq 0. \qquad (j \in S,\ k \in K_j)$$

Problem (P_T) is now a standard linear programming problem. In the following discussion, we take advantage of the special structure of this problem to prove that its solution has several interesting properties. We shall then use these properties to show that there is always an optimal solution for this problem and that for any basic optimal solution there is a corresponding optimal solution to problem (P_0).

We shall derive a stationary strategy as a function that for each $i \in S$ selects exactly one variable x_i^k where $k \in K_i$. This definition of stationary strategies will coincide with that of the preceding section. For any stationary strategy, we shall now prove the following theorem which has been given by De Ghellinck and Eppen [17].

<u>Theorem 6. 8.</u> If the system of equations (6.15) is restricted to the variables x_i^k selected by any stationary strategy then:

(i) The corresponding subsystem has a unique solution.

(ii) If $a_j \geq 0$ ($j \in S$) then $x_j^k \geq 0$ ($j \in S$).

(iii) If $a_j > 0$ ($j \in S$) then $x_j^k > 0$ ($j \in S$).

We shall now use Theorem 6. 8 to prove an important relationship between stationary strategies and the basic feasible solutions of (6.15) when $a_j > 0$ ($j \in S$). The following two theorems have also been given by De Ghellinck and Eppen [17].

Theorem 6. 9. Whenever $a_j > 0$ ($j \in S$), there exists a one-to-one correspondence between stationary strategies and basic feasible solutions of (6.15). Moreover, any basic feasible solution is nondegenerate.

Proof. Theorem 6. 8 actually states that when $a_j > 0$ ($j \in S$), a stationary strategy has a corresponding unique solution of (6.17) which has N positive variables. This is, by definition, a basic nondegenerate feasible solution of (6.15). Conversely, if x_i^k is a feasible solution of (6.15), we have

$$(6.17) \quad \sum_{k \in K_j} x_j^k = a_j + \beta \sum_{i \in S} \sum_{k \in K_i} p_{ij}^k x_i^k \geq a_j > 0, \quad (j \in S)$$

and thus, at least one variable x_j^k has to be positive; hence, there is exactly one $x_i^k > 0$ for each state j. This uniquely defines a stationary strategy. These complete the proof.

Any strategy which is associated in this way with an optimal basic solution will be called an optimal stationary strategy. In the next theorem we shall show that although an optimal stationary strategy is derived with a specific set of values $a_j > 0$, this strategy remains optimal for any nonnegative right-hand side a_j.

Theorem 6. 10. Whenever the right-hand side a_j of

(6.15) is strictly positive, the problem (P_T) has an optimal basic solution and its dual has a unique optimal solution. Any optimal stationary strategy associated with it remains optimal for any nonnegative right-hand side of a_j.

Proof. From Theorem 6. 8 it is clear that there are feasible solutions, and we also know from Lemma 6. 7 that the objective function is bounded. This guarantees the existence of an optimal solution for both the problem and its dual. By Theorem 6. 9 any basic solution will be nondegenerate and by the complementary slackness conditions, any optimal solution of the dual should satisfy the corresponding system of N dual equalities (which is nonsingular), and therefore the dual solution is unique. The discussion of the dual problem will be stated in detail in the next section.

To prove the second part of the theorem, we note that the optimality of a given basic feasible solution of a linear programming problem depends on the objective function and not on the right-hand side. Changing the latter can only affect the feasibility for any nonnegative value of a_j. These complete the proof.

The following corollary is useful in the next section.

Corollary 6. 11. For all positive right-hand side

$a_j > 0$ (say $a_j = 1/N$), there exists any basic solution with property such that for each $i \in S$, there is only one k such that $x_i^k > 0$ and $x_i^k = 0$ for k otherwise.

Proof. This is a direct consequence of Theorems 6. 9 and 6. 10.

From a series of the theorems, we know the special structures of this linear programming problem. Then under the suitable initial distribution (say $a_j = 1/N$), we can get an optimal basic solution which corresponds to an optimal stationary strategy. We have seen two algorithms; one is the Policy Iteration Algorithm discussed in the preceding section and another is a Linear Programming Algorithm discussed in this section. However, we cannot answer which of these is more advantage to compute an optimal strategy. In the next section we shall discuss a relation between the two algorithms.

6. 4. Relation between the Two Algorithms

In section 6. 2 we have derived the Policy Iteration Algorithm, which may be considered to be a successive approximation method in policy space of Dynamic Programming. In Section 6. 3 we have formulated the same problem by Linear Programming. In this section we shall show that these two algorithms are equivalent in mathematical programming.

We now rewrite a linear programming problem discussed in the preceding section as the primal problem.

Primal problem:

$$\text{(6.18)} \qquad \text{Max} \sum_{j \in S} \sum_{k \in K_j} r_j^k x_j^k$$

subject to

$$\text{(6.19)} \qquad \sum_{k \in K_j} x_j^k + \beta \sum_{i \in S} \sum_{k \in K_i} p_{ij}^k x_i^k = a_j, \qquad (j \in S)$$

$$\text{(6.20)} \qquad x_j^k \geq 0. \qquad (j \in S,\ k \in K_j)$$

The dual problem of the above one is significant as discussed in the preceding section.

Dual problem:

$$\text{(6.21)} \qquad \text{Min} \sum_{i \in S} a_i v_i$$

subject to

$$\text{(6.22)} \qquad v_i \geq r_i^k + \beta \sum_{j \in S} p_{ij}^k v_j, \qquad (i \in S,\ k \in K_i)$$

(6.23) $\quad v_i$; unconstrained in sign for $i \in S$.

This dual problem can be immediately derived from the discussion of Section 6. 2 as follows. For an optimal stationary strategy $\pi = f^{\infty}$, we have

Dual \ Primal	$x_1^1 \geq 0$	$x_1^2 \geq 0$	$\cdots$	$x_2^1 \geq 0$	$x_2^2 \geq 0$	$\cdots$	$x_N^1 \geq 0$	$x_N^2 \geq 0$	$\cdots$	Relations	Variables
v_1	$1-\beta p_{11}^1$	$1-\beta p_{11}^2$	$\cdots$	$-\beta p_{21}^1$	$-\beta p_{21}^2$	$\cdots$	$-\beta p_{N1}^1$	$-\beta p_{N1}^2$	$\cdots$	$=$	a_1
v_2	$-\beta p_{12}^1$	$-\beta p_{12}^2$	$\cdots$	$1-\beta p_{22}^1$	$1-\beta p_{22}^2$	$\cdots$	$-\beta p_{N2}^1$	$-\beta p_{N2}^2$	$\cdots$	$=$	a_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$			$\vdots$	$\vdots$		$\vdots$	$\vdots$
v_{N-1}	$-\beta p_{1,N-1}^1$	$-\beta p_{1,N-1}^2$	$\cdots$	$-\beta p_{2,N-1}^1$	$-\beta p_{2,N-1}^2$	$\cdots$	$-\beta p_{N,N-1}^1$	$-\beta p_{N,N-1}^2$	$\cdots$	$=$	a_{N-1}
v_N	$-\beta p_{1N}^1$	$-\beta p_{1N}^2$	$\cdots$	$-\beta p_{2N}^1$	$-\beta p_{2N}^2$	$\cdots$	$1-\beta p_{NN}^1$	$1-\beta p_{NN}^2$	$\cdots$	$=$	a_N
Relations	$\geq$	$\geq$	$\cdots$	$\geq$	$\geq$	$\cdots$	$\geq$	$\geq$	$\cdots$		
Constants	r_1^1	r_1^2	$\cdots$	r_2^1	r_2^2	$\cdots$	r_N^1	r_N^2	$\cdots$		

Fig. 6. 1.   The Tucker diagram for the Markovian decision process with discounting

(6.24) $$V_\beta(f^\infty) \geq L(g)\, V_\beta(f^\infty) = r(g) + \beta P(g)\, V_\beta(f^\infty)$$

for all $g \in F$. Writing the above equation elementwise, we have the constraints (6.22). While, the objective function is the discounted total expected return starting in the initial distribution a;

(6.25) $$a V_\beta(f^\infty) = \sum_{i \in S} a_i v_i ,$$

which implies the dual problem. It is more comprehensive to write the Tucker Diagram in Fig. 6. 1 for the primal and dual problems.

We shall consider to solve the primal problem. Since the constraints of the problem are equalities, we must use Two Phase Method or Composite Algorithms to obtain an initial basic feasible solution. But, Corollary 6. 11 implies that each $j \in S$ there is one such that $x_j^k > 0$ and $x_j^k = 0$ otherwise. So, we omit phase I and can obtain a basic feasible solution. In order to obtain an initial basic feasible solution, we use the usual simplex criterion for each $j \in S$. For example, we may apply

(6.26) $$-r_j^* = \min_{k \in K_j} [-r_j^k], \qquad (j \in S)$$

or

(6.27) $$r_j^* = \max_{k \in K_j} \; r_j^k . \qquad (j \in S)$$

which corresponds to an initial strategy discussed in Section 6. 2. We note that the asterisk stands for the data of a basic solution throughout this section.

For any basic feasible solution, we have the basic matrix;

$$
(6.28) \qquad B = \begin{pmatrix} 1-\beta p_{11}^* & -\beta p_{21}^* & \cdots & -\beta p_{N1}^* \\ -\beta p_{12}^* & 1-\beta p_{22}^* & \cdots & -\beta p_{N2}^* \\ \vdots & \vdots & & \vdots \\ -\beta p_{1N}^* & -\beta p_{2N}^* & \cdots & 1-\beta p_{NN}^* \end{pmatrix},
$$

where p_{ji}^* is the transition probability according to the basic solution. Adding the row vector corresponding to the coefficients of the objective function, we have the modified basic matrix;

$$
(6.29) \qquad \bar{B} = \begin{pmatrix} 1 & -r_1^* & -r_2^* & \cdots & -r_N^* \\ 0 & 1-\beta p_{11}^* & -\beta p_{21}^* & \cdots & -\beta p_{N1}^* \\ 0 & -\beta p_{12}^* & 1-\beta p_{22}^* & \cdots & -\beta p_{N2}^* \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & -\beta p_{1N}^* & -\beta p_{2N}^* & \cdots & 1-\beta p_{NN}^* \end{pmatrix},
$$

where we add the $(N + 1) \times 1$ unit vector with the first element unity to the first column. From the preceding section, B is nonsingular (Theorem 6. 8), so is $\overline{B}$. Thus, let $\overline{B}^{-1}$ be the inverse matrix of $\overline{B}$. We have

$$\overline{B}^{-1} = \begin{bmatrix} 1 & \mu \\ 0 & B^{-1} \end{bmatrix}. \tag{6.30}$$

It is clear that $\overline{B}\,\overline{B}^{-1} = I$ (identity matrix). So, we have

$$\mu = r^{*T} B^{-1}, \tag{6.31}$$

or

$$B^T \mu^T = r^*, \tag{6.32}$$

where the superscript T denotes the transpose of the matrix. From its definition (6.30), μ is the vector of the simplex multipliers for the primal problem, and that μ is also the vector of the dual variables. Thus, we have $\mu = (v_1, v_2, \ldots, v_N)$. Rewriting (6.32) elementwise, we have

$$v_i = r_i^* + \beta \sum_{j \in S} p_{ij}^* v_j, \qquad (i \in S) \tag{6.33}$$

which corresponds to Value Determination Operation in the Policy Iteration Algorithm discussed in Section 6. 2. In other words, solving the system of N linear equations (6.33) is to find the simplex multipliers (and also the dual variables) for the primal problem. (This derivation is straightfoward from the complementary slackness principle.)

Next, we consider the simplex criterion for the next step using these simplex multipliers. Let the simplex multiplier for the objective function be 1 and we write

$$(6.34) \qquad \mu = (1, v_1, v_2, \ldots, v_N).$$

The coefficients of the corresponding linear programming problem can be written by

$$(6.35) \qquad A = \begin{pmatrix} -r_1^1 & -r_1^2 & \cdots\cdots & -r_N^1 & \cdots & -r_N^{k_N} \\ 1-\beta p_{11}^1 & 1-\beta p_{11}^2 & \cdots\cdots & -\beta p_{N1}^1 & \cdots & -\beta p_{N1}^{k_N} \\ -\beta p_{12}^1 & -\beta p_{12}^2 & \cdots\cdots & -\beta p_{N2}^1 & \cdots & -\beta p_{N2}^{k_N} \\ \vdots & \vdots & & \vdots & & \vdots \\ -\beta p_{1N}^1 & -\beta p_{1N}^2 & \cdots\cdots & 1-\beta p_{NN}^1 & \cdots & 1-\beta p_{NN}^{k_N} \end{pmatrix}.$$

The simplex criterion of the next step is;

$$(6.36) \qquad [\Delta_i^k] = \mu A$$
$$= [-r_i^k - \beta \sum_{j \in S} p_{ij}^k v_j + v_i].$$

It is evident that for the basic variables

(6.37) $$\Delta_i^* = -r_i^* - \beta \sum_{j\in S} P_{ij}^* v_j + v_i = 0 \qquad (i \in S)$$

If for all $i \in S$ and $k \in K_i$,

(6.38) $$\Delta_i^k = -r_i^k - \beta \sum_{j\in S} P_{ij}^k v_j + v_i \geq 0,$$

or by using (6.37), for all $i \in S$ and $k \in K_i$

(6.39) $$r_i^* + \beta \sum_{j\in S} P_{ij}^* v_j \geq r_i^k + \beta \sum_{j\in S} P_{ij}^k v_j,$$

we have an optimal solution from the basic theory of linear programming [15].

While, if there is at least one pair $i \in S$ and $k \in K_i$ such that

(6.40) $$\Delta_i^k = -r_i^k - \beta \sum_{j\in S} P_{ij}^k v_j + v_i < 0,$$

or by using (2.38)

(6.41) $$r_i^* + \beta \sum_{j\in S} P_{ij}^* v_j < r_i^k + \beta \sum_{j\in S} P_{ij}^k v_j,$$

there exists an improved solution, or an improved strategy, which corresponds to <u>Policy Improvement Routine</u> in the Policy Iteration Algorithm.

Consequently, Policy Iteration Algorithm is only a special extension of linear programming such that pivot operations for many (at most N) variables are performed simultaneously. We have already seen in Section 6. 2 that the substitutions for many variables imply an improved strategy. But, even if there is only one pair $i \in S$ and $k \in K_i$ satisfying (6.41), we must solve the system of N linear equations (6.33). It is a disadvantage to perform its computation. These questions arising in computing an optimal strategy will be discussed in Section 7. 5.

6. 5. Return Structures

In the preceding discussion we assume that we receive the return when the system is in state i .

We extend the return structures as follows:

When the system is in state i , two things happen:

(i) We receive the return r_i'.

(ii) If the system moves to state j at next time, we receive the return r_{ij}'.

Under the above return structures, $v_i(n)$, the n step total expected return is;

$$(6.42) \qquad v_i(n) = r_i' + \sum_{j \in S} p_{ij} \, [r_{ij}' + \beta v_j(n-1)]$$

$$= r_i' + \sum p_{ij} r_{ij}' + \beta \sum_{j \in S} p_{ij} v_j(n-1).$$

Substituting

$$(6.43) \qquad r_i = r_i' + \sum_{j \in S} p_{ij} r_{ij}' ,$$

we have the same return structures treated in the preceding discussion. Thus, in the sequel we consider only the return r_i^k for the discrete time models.

6. 6. Examples

This section gives two numerical examples and their solutions by using two algorithms, i.e., the policy iteration algorithm and a linear programming algorithm.

Example 1. The first example has been given in the Introduction. Now we define $F = \{f_1, \ f_2\}$, where $f_k(1) = 1, \quad f_k(2) = k$.
Then

$$P(f_1) = \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}, \qquad r(f_1) = \begin{bmatrix} 3 \\ -2 \end{bmatrix},$$

$$P(f_2) = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}, \qquad r(f_2) = \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

where f_1 denotes a strategy of a rapid repair and f_2 a strategy of a usual repair. We now shall solve this problem setting a discount factor $\beta = 0.9$.

First we apply the policy iteration algorithm. Take an initial strategy f_1. From Value Determination Operation we have

$$v_1 = 1380/91 \qquad v_2 = 880/91$$

By using the values obtained above and Policy Improvement

Routine, we obtain

$$r_2^2 + \beta \sum_{j=1}^{2} p_{ij}^2 v_j = 881/91 > 880/91 = v_2,$$

which implies an improved strategy f_2. Returning to Value Determination Operation we have

$$v_1 = 1110/73 \qquad v_2 = 710/73$$

which concludes with a 0.9-optimal f_2.

Second we apply the linear programming approach. Under an initial distribution

$$a = (\ 0.5,\ 0.5\ ),$$

we have the following linear programming problem:

$$\text{Max} \quad 3x_1^1 - 2x_2^1 - x_2^2$$

subject to

$$0.37\,x_1^1 - 0.54\,x_2^1 - 0.36\,x_2^2 = 0.5$$

$$-0.27\,x_1^1 + 0.64\,x_2^1 + 0.46\,x_2^2 = 0.5$$

$$x_1^1,\ x_2^1,\ x_2^2 \geq 0$$

Thus the optimal solution is

$$x_1^1 = 410/73 \qquad x_2^2 = 320/73$$

and the objective function is 910/73. The value of our objective function coincides with

$$a V_\beta(f_2^\infty) = [\ 0.5 \quad 0.5\ ] \begin{bmatrix} 1110/73 \\ \\ 710/73 \end{bmatrix} = 910/73$$

Example 2. (Howard's Taxicab Problem [40, p. 44]). Consider the problem of a taxicab driver whose territory encompasses three towns, A, B, and C. If he is in town A, he has three actions:

1. He can cruise in the hope of picking up a passenger by being hailed.

2. He can drive to the nearest cab stand and wait in line.

3. He can pull over and wait for a radio car.

If he is in town C, he has the same three actions, but if he is in town B, the last action is not present because there is no radio cab service in that town. For a given town and given action, there is a probability that the next trip will go to each on the towns A, B, and C and a corresponding return in monetary units associated with each such trip. This return presents the income from the trip after all necessary expences have deducted. For example, in the case of actions 1 and 2, the cost of cruising and of driving to the nearest stand must be included in calculating the returns. The probabilities of transition and the returns depend upon the action because different customer polulation will be encountered under each action.

If we identify being in towns A, B, and C with states 1, 2, and 3, respectively, then we have Table 6. 2.

Here we assume $\beta = 0.90$. Let an initial policy

Table 6. 2. Data for the taxicab problem. (From R. A. Howard, Dynamic programming and Markov Processes, M.I.T. Press, Cambridge, 1960.)

State	Action	Probability			Return			
i	k	p_{i1}^k	p_{i2}^k	p_{i3}^k	r_{i1}^k	r_{i2}^k	r_{i3}^k	$r_i^k = \sum_{j=1}^{3} p_{ij}^k r_{ij}^k$
1	1	1/2	1/4	1/4	10	4	8	8
	2	1/16	3/4	3/16	8	2	4	2.75
	3	1/4	1/8	5/8	4	6	4	4.25
2	1	1/2	3/4	3/16	4	6	4	16
	2	1/4	1/8	5/8	14	0	18	15
3	1	1/4	1/4	1/2	10	2	8	7
	2	1/8	3/4	1/8	6	4	2	4
	3	3/4	1/16	3/16	4	0	8	4.5

v_1		91.3		119.4		121.7	
v_2		97.6		134.5		135.3	
v_3		90.0		121.9		122.8	
	V.D.O.		P.I.V / V.D.O.		P.I.V. / V.D.O.		P.I.V.
f	1		1		2		2
	1		2		2		2 STOP
	1		2		2		2

V.D.O.: Value Determination Operation

P.I.V.: Policy Improvement Routine

Table 6. 3. Solution of the taxicab problem by the Policy Iteration Algorithm with $\beta = 0.9$.

be

$$f = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

which is derived from $\max_{k \in K_i} r_i^k$ for each $i \in S$.

Policy Iteration Algorithm yields an optimal strategy. The calculations of Policy Iteration Algorithm are given in Table 6. 3.

We can also solve this problem by linear programming. Here we omit the solution.

CHAPTER VII

MARKOVIAN DECISION PROCESSES WITH NO DISCOUNTING

7. 1. Introduction

We shall consider Markovian decision processes with no discounting in this chapter. We shall treat only the process over an infinite planning horizon. In this process, the total expected return will be usually divergent. (In Section 7. 6 we consider a special case in which the total expected return is finite.)

So, we shall not expect the elegant discussion as we have done in the preceding chapter since Markov chain under consideration may change its state classification structure from strategy to strategy.

First, we shall discuss the special structure processes, i.e., the completely ergodic process and the terminating process (see the definitions below). Second, we shall discuss the general process that the state classification may change from strategy to strategy.

Two approaches have been considered for Markovian decision processes with no discounting. One is treating $\beta = 1$ as a limiting case of $\beta < 1$, where we can use the discussion of the preceding chapter. Anoter is treating directly the long-run average return per unit time. We shall mainly treat the first approach, but we shall partially treat the second approach.

7. 2. Policy Iteration Algorithm

This section provides the preliminaries on the theory of Markov chains. This section further states the policy iteration algorithm for the completely ergodic process, which has been given by Howard [40].

First, we shall show the well-known fact of matrices.

Lemma 7. 1. For any $N \times N$ matrix A, if A^n tends to 0 (zero matrix) as n tends to infinity, then $(I - A)$ is nonsingular, and

$$(7.1) \qquad (I - A)^{-1} = \sum_{i=0}^{\infty} A^i .$$

The proof of this lemma will be found, e.g., in Kemeny and Snell [47, p. 22].

Next we shall show

Lemma 7. 2. For any $N \times N$ Markov matrix P, if $\sum_{i=0}^{n-1} P^i / n$ tends to P^* as n tends to infinity, then $(1 - \beta) \sum_{i=0}^{\infty} \beta^i P^i$ tends to P^* as $\beta \to 1 - 0$.

Proof. The hypothesis asserts Césaro summability of $\{P^n\}$ to P^*. Thus Césaro summability of $\{P^n\}$ to P^* implies Abel summability of $\{P^n\}$ to P^*, i.e.,

$$(7.2) \qquad (1-\beta) \sum_{i=0}^{\infty} \beta^i P^i \longrightarrow P^* \qquad \text{as} \quad \beta \to 1-0,$$

which completes the proof.

The next lemma will be important for the latter discussion. The proof of the lemma is found in Blackwell [9] and others.

Lemma 7. 3. Let P be any $N \times N$ Markov matrix.

(a) The sequence $\sum_{i=0}^{n-1} P^i / n$ converges as $n \to \infty$ to a Markov matrix P^* such that

(7.3) $$P P^* = P^* P = P^* P^* = P^*.$$

(b) $I - (P - P^*)$ is nonsingular, and

(7.4) $$H(\beta) = \sum_{i=0}^{\infty} \beta^i (P^i - P^*) \longrightarrow H = (I - P + P^*)^{-1} - P^*$$

as $\beta \longrightarrow 1 - 0$.

(7.5) $$H(\beta) P^* = P^* H(\beta) = H P^* = P^* H = 0,$$

and

(7.6) $$(I - P) H = H (I - P) = I - P^*.$$

(7.7) (c) rank $(I - P)$ + rank $P^* = N$.

(d) For every $N \times 1$ column vector c, the system

(7.8) $$Px = x , \qquad P^* x = P^* c$$

has a unique solution x.

Following the notation of the preceding chapter, the total expected return up to time t starting in each state is

$$(7.9) \qquad \sum_{i=0}^{n-1} P_i(\pi)\, r(f_{i+1})$$

for any strategy $\pi = (f_1, f_2, \ldots)$. Thus the limit infimum of the average return per unit time as n tends to infinity is

$$(7.10) \qquad \Gamma(\pi) = \liminf_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} P_i(\pi)\, r(f_{i+1}).$$

Our problem is then to find a strategy which maximizes (7.10) under all strategies π, i.e., an optimal strategy π^* such that

$$(7.11) \qquad \Gamma(\pi^*) \geq \Gamma(\pi)$$

for all π.

Derman [20] has shown that there is an optimal strategy which is stationary, where an optimal strategy is considered under the average return criterion. The same result has also been proved by Blackwell [9] and others.

<u>Theorem 7. 4.</u> There is an optimal strategy which is stationary.

This theorem shows that there is an optimal stationary strategy under the average criterion, where an optimal strategy is attained simultaneously for all initial states. Thus, an optimal strategy remains optimal for any initial distribution a.

In the sequel (up to Section 7. 5), we shall discuss a special structure problem, i.e., the so-called completely ergodic process.

Definition 7. 5. For a Markovian decision process, Markov chain under consideration is always ergodic whatever strategies we choose. Then the process is called the completely ergodic process.

Examples presented in Section 6. 6 are all completely ergodic processes. Though there really are many examples which are not completely ergodic, we shall discuss only the completely ergodic processes up to Section 7. 5. The general case that there are several ergodic sets plus some transient states will be shown in Sections 7. 7 and 7. 8.

Note that the process, which extends the range of decisions to include randomized strategies, is also completely ergodic from its definition.

For any stationary strategy f^∞ of the completely ergodic process, we have

$$(7.12)\ \liminf_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P_i(\pi)r(f_{i+1})=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^i(f)r(f)=P^*(f)r(f),$$

where P^* is the limiting matrix which is composed of all identical rows $\pi(f) = [\pi_j(f)]$. That is,

$$P^*(f) = \mathbb{1}\,\pi(f), \tag{7.13}$$

where $\mathbb{1}$ is the $N \times 1$ column vector with all elements 1. Then, the $N \times 1$ row vector $\pi(f)$ is a unique solution of

$$\pi(f) = \pi(f) P(f), \tag{7.14}$$

$$\pi_j(f) > 0, \qquad (j \in S) \tag{7.15}$$

$$\pi(f)\,\mathbb{1} = 1. \tag{7.16}$$

Thus an average return per unit time for a stationary strategy starting in each state is

$$P^*(f) r(f) = \mathbb{1}\,\pi(f) r(f) = \left(\sum_{j \in S} \pi_j(f) r_j^k\right) \mathbb{1}, \tag{7.17}$$

where $k = f(j)$. This means that the average return per unit time is the weighted sum of r_j^k over the limiting probabilities $\pi_j(f)$, and is identical for any initial state $i \in S$. That is, the average return is independent of any initial distribution.

We have known from the preceding discussion that our problem is to find an optimal strategy within the finite set of stationary strategies because there is an optimal stationary strategy. That is, our problem is a combinatorial one and we can apply the direct enumeration method. (This fact is valid not only for the completely ergodic process but also for the general process in Sections 7. 7 and 7. 8. In most cases, however, the direct enumeration approach seems to be impossible because the number of stationary stratigies, $K_1 \times K_2 \times \cdots \times K_N$, is too numerous to compute each quantity. Thus we shall need efficient algorithms to find an optimal strategy.

For any stationary strategy f^∞, $V^n(f)$, the total expected return vector up to time n, satisfies the following recursive relation:

$$(7.19) \qquad V^n(f) = r(f) + P(f)\,V^{n-1}(f), \qquad (n = 1, 2, \ldots)$$

where

$$(7.20) \qquad V^0(f) = 0.$$

Throughout this chapter, we restrict our attention to stationary strategies, then we may write a stationary strategy f instead of f^∞. We have the following

Lemma 7. 6. If $P^n(f)$ tends to $P^*(f)$ as n tends to infinity, i.e., $P(f)$ is regular, then we have

$$(7.21) \qquad V^n(f) = n\,g\mathbf{1} + v(f) + \varepsilon(n, f),$$

where $\varepsilon(n, f)$ tends to zero as n tends to infinity, $g = \pi(f)\,r(f)$, and $v(f) = H(f)\,r(f)$.

The proof of the above lemma is easily obtained by using Lemma 7. 3. Using the above lemmas, we have the policy iteration algorithm for the completely ergodic process. Here we shall only give the algorithm. The heuristic derivation of the algorithm and the proof of the convergence to an optimal strategy are found in Howard [40].

Value Determination Operation

Take any stationary f^∞. Solve

$$g + v_i = r_i^k + \sum_{j=1}^{N-1} p_{ij}^k v_j \quad .$$

for g, v_1, v_2, ..., v_{N-1} (setting $v_N = 0$), where the superscript k is determined by the chosen strategy f^∞.

Policy Improvement Routine

Using the values v_i and g, find the element of $G(i, f)$ for each $i \in S$ such that

$$r_i^k + \sum_{j=1}^{N-1} p_{ij}^k v_j > g + v_i$$

for all $k \in K_i$. If $G(i, f)$ is empty for all $i \in S$, f^∞ is optimal and g is the average return per unit time, v_1, ..., v_{N-1} are the relative bias terms. If at least $g(i) \in G(i, f)$ for some i, make an improved strategy g^∞ such that $g(i) \in G(i, f)$ for some i and $g(i) = f(i)$ for $G(i, f)$ empty, and return to Value Determination Operation.

In Policy Improvement Routine, if there are two or more actions satisfying $G(i, f)$ for some $i \in S$, we should apply an improved strategy $g(i)$ such that

$$\max_{k \in K_i} [r_i^k + \sum_{j=1}^{N-1} p_{ij}^k v_j]. \tag{7.22}$$

The reason of this choice will be presented in Section 7. 4 from the viewpoint of linear programming.

7. 3. Linear Programming Algorithm

We shall show that the completely ergodic Markovian decision process is also formulated by a linear programming problem. For any stationary strategy of the completely ergodic process, the average return per unit time is

$$g(f) = \pi(f) r(f), \tag{7.23}$$

where $\pi(f)$ is the limiting vector of $P(f)$, which saisfies (7.14)-(7.16). Thus our problem is to find an optimal strategy f^* such that

$$\pi(f^*) r(f^*) = \max_{f \in F} \pi(f) r(f). \tag{7.24}$$

It is more convenient to extend the range of decisions to include randomized strategies. Note that the assumption of complete ergodicity also holds for the extension to randomized strategies. Then, let d_j^k ($j \in S$, $k \in K_j$) be the joint probability that the system is in state i and the decision k is make, where d_j^k is independent of time n because we restrict ourselves to stationary strategies. It is evident that

(7.25) $\sum_{k \in K_j} d_j^k = 1, \qquad 0 \le d_j^k \ (\le 1). \qquad (j \in S,\ k \in K_j)$

Then the objective function of our problem is

(7.26) $\sum_{j \in S} \sum_{k \in K_j} \pi_j(f) r_j^k d_j^k$

from (7.24) and the definition of d_j^k. While, the constraints using d_j^k are

(7.27) $\pi_j(f) - \sum_{i \in S} \sum_{k \in K_i} \pi_i(f) p_{ij}^k d_i^k = 0, \qquad (j \in S)$

(7.28) $\sum_{j \in S} \pi_j(f) = 1,$

(7.29) $\pi_j(f) > 0 \qquad (j \in S)$

Setting

(7.30) $x_j^k = \pi_j(f) d_j^k \ge 0$

and using the fact that $\pi_j(f) = \sum_{k \in K_j} x_j^k$ for $j \in S$, we have the following linear programming problem:

(7.31) Max $\sum_{j \in S} \sum_{k \in K_j} r_j^k x_j^k$

subject to

$$(7.32) \qquad \sum_{k\in K_j} x_j^k - \sum_{i\in S}\sum_{k\in K_i} p_{ij}^k x_i^k = 0, \qquad (j \in S)$$

$$(7.33) \qquad \sum_{j\in S}\sum_{k\in K_j} x_j^k = 1,$$

$$(7.34) \qquad x_j^k \geq 0. \qquad (j\in S,\ k\in K_j)$$

For the completely ergodic process, rank($I - P$) $= N - 1$ because rank $P^* = 1$. So, one of the constraints (7.32) is redundant. We omit one constraint, e.g., for $j = N$ of (7.32). Then the constraints are (7.32) for $j = 1, 2, \ldots, N - 1$, and (7.33).

Theorem 7. 7. There exists any basic feasible solution with property such that for each $i \in S$, there is only one k such that $x_i^k > 0$ and $x_i^k = 0$ for k otherwise.

Proof. Our linear programming problem has N constraints, where one redundant constraint is omitted. Since rank of the constraints is N, there are N positive variables x_i^k and zero otherwise for any basic feasible solution from the basic properties of linear programming. For the coefficients of the constraints, $-p_{ij}$($i \neq j$) is nonpositive, ($1 - p_{ii}^k$) is positive and x_i^k is nonnegative, so there is at least one term

$(1 - p_{ii}^k)$ in which x_i^k is positive for each i. That is, for each i there is at least one $x_i^k > 0$. If there are two $x_j^k > 0$ for any one j, there is some i without the term $(1 - p_{ii}^k)$ somewhere because any basic feasible solution has N positive variables x_j^k and zero otherwise. This contradicts that the right-hand side is zero or unity. Therefore, for each i there is only one $x_i^k > 0$ and zero otherwise, which completes the proof.

Corollary 7. 8. Any basic feasible solution of the linear programming problem (7.31)-(7.34) yields a pure stationary strategy.

Proof. From (7.30) and $\pi_j(f) = \sum_{k \in K_j} x_j^k$, we have

$$d_j^k = x_j^k \Big/ \sum_{k \in K_j} x_j^k . \tag{7.35}$$

Thus $d_j^k = 0$ or 1 from Theorem 7. 7, which completes the proof.

From the above discussion, an optimal solution of the linear programming problem (7.31)-(7.34) yields an optimal pure stationary strategy f^∞. Also the primal variable $x_j^k > 0$ gives the limiting probability $\pi_j(f) > 0$ for state j.

7. 4. Relation between the Two Algorithms

We have shown in Section 6. 4 that Policy Iteration and Linear Programming Algorithms are equivalent in mathematical programming for the discounted Markovian decision process. In this section we shall also show the similar relation for the completely ergodic process.

We now rewrite a linear programming problem discussed in the preceding section as a primal problem.

Primal Problem:

(7.36) $$\text{Max} \quad \sum_{j \in S} \sum_{k \in K_j} r_j^k x_j^k$$

subject to

(7.37) $$\sum_{k \in K_j} x_j^k - \sum_{i \in S} \sum_{k \in K_i} p_{ij}^k x_i^k = 0, \qquad (j = 1, 2, \ldots, N-1)$$

(7.38) $$\sum_{j \in S} \sum_{k \in K_j} x_j^k = 1,$$

(7.39) $$x_j^k \geq 0. \qquad (j \in S,\ k \in K_j)$$

Here we qmit a redundant constraint for $j = N$ in (7.37).

The dual problem of the above one is significant since any basic feasible solution corresponds to the N dual equalities (which is nonsingular) from Theorem 7. 7 and its proof. Let dual variables be (v_1, v_2, ..., v_N).

Dual Problem:

(7.40) $$\text{Min} \quad v_N$$

subject to

(7.41) $$v_N + v_i \geq r_i^k + \sum_{j=1}^{N-1} p_{ij}^k v_j, \qquad (i \in S,\ k \in K_i)$$

(7.42) $\quad v_i$; unconstrained in sign for $i \in S$.

Since the primal problem has an optimal solution and its value is g, we may write the dual problem using duality theorem (see, e.g., Dantzig [15]) as follows.

Dual Problem:

(7.43) $\quad$ Max $\quad g$

subject to

(7.44) $$g + v_i \geq r_i^k + \sum_{j=1}^{N-1} p_{ij}^k v_j, \qquad (i \in S,\ k \in K_i)$$

(7.45) $\quad g,\ v_i$; unconstraintd in sign for $i = 1, 2, \ldots, N - 1$.

where we suppose $v_N = 0$ in (7.44).

The dual problem is immediately derived from the discussion of Section 7. 2. The method of deriving the dual problem is almost similar to that of Section 6. 2. So, we omit here the derivation.

It is more comprehensive to write the Tucker Diagram in Fig. 7. 1 for primal and dual problems. In the dual problem, setting $v_N = 0$ corresponds to omitting a redundant constraint for $j = N$ in (7.37) of the primal problem.

As the similar discussion of Section 6. 4, we can omit Phase I and can immediately obtain a basic feasible solution using Theorem 7. 7. As an initial basic variable, we may, for example, apply

(7.46) $$-r_j^* = \min_{k \in K_j} [-r_j^k]. \qquad (j \in S)$$

Dual	Primal									
		$x_1^1 \geq 0$	$x_1^2 \geq 0$	$\cdots$	$x_2^1 \geq 0$	$x_2^2 \geq 0$	$\cdots$	$x_N^1 \geq 0$	$x_N^2 \geq 0 \cdots$	Relations	Variables
	v_1	$1-p_{11}^1$	$1-p_{11}^2$	$\cdots$	$-p_{21}^1$	$-p_{21}^2$	$\cdots$	$-p_{N1}^1$	$-p_{N1}^2 \cdots$	$=$	0
	v_2	$-p_{12}^1$	$-p_{12}^2$	$\cdots$	$1-p_{22}^1$	$1-p_{22}^2$	$\cdots$	$-p_{N2}^1$	$-p_{N2}^2 \cdots$	$=$	0
	$\cdot$	$\cdot$	$\cdot$		$\cdot$	$\cdot$		$\cdot$	$\cdot$	$\cdot$	$\cdot$
	$\cdot$	$\cdot$	$\cdot$		$\cdot$	$\cdot$		$\cdot$	$\cdot$	$\cdot$	$\cdot$
	$\cdot$	$\cdot$	$\cdot$		$\cdot$	$\cdot$		$\cdot$	$\cdot$	$\cdot$	$\cdot$
	v_{N-1}	$-p_{1,N-1}^1$	$-p_{1,N-1}^2$	$\cdots$	$-p_{2,N-1}^1$	$-p_{2,N-1}^2$	$\cdots$	$-p_{N,N-1}^1$	$-p_{N,N-1}^2 \cdots$	$=$	0
	g	1	1	$\cdots$	1	1	$\cdots$	1	$1 \cdots$	$=$	1
	Relations	$\geq$	$\geq$	$\cdots$	$\geq$	$\geq$	$\cdots$	$\geq$	$\geq \cdots$		
	Constants	r_1^1	r_1^2	$\cdots$	r_2^1	r_2^2	$\cdots$	r_N^1	$r_N^2 \cdots$		

Fig. 7. 1. The Tucker diagram for the completely ergodic Markovian decision process.

For any basic feasible solution, we have the basic matrix:

(7.47) $$B = \begin{pmatrix} 1-p_{11}^* & -p_{21}^* & \cdots & \cdots & -p_{N1}^* \\ -p_{12}^* & 1-p_{22}^* & \cdots & \cdots & -p_{N2}^* \\ \vdots & \vdots & & & \vdots \\ -p_{1,N-1}^* & -p_{2,N-1}^* & \cdots & \cdots & -p_{N,N-1}^* \\ 1 & 1 & \cdots & \cdots & 1 \end{pmatrix},$$

where p_{ji}^* is the transition probability corresponding to a basic solution. From the similar discussion of Section 6. 4, defining the simplex multipliers (also dual variables) $\mu = (\ v_1,\ \ v_2,\ \ldots,\ v_{N-1},\ \ g\ )$, we have

(7.48) $$B^T\mu^T = r^*,$$

where r^* is the $N \times 1$ row vector whose ith element is r_i^*, and the superscript T denotes the transpose. Rewriting (7.48) elementwise, and supposing $v_N = 0$, we have

(7.49) $$g + v_i = r_i^* + \sum_{j=1}^{N-1} p_{ij}^* v_j\ , \qquad (i \in S)$$

which corresponds to Value Determination Operation in the Policy Iteration Algorithm discussed in Section 7. 2.

The simplex criterion of the next step is

(7.50) $$[\Delta_i^k] = [-r_i^k - \sum_{j=1}^{N-1} p_{ij}^k v_j + g + v_i].$$

It is evident that for the basic variables, we have

$$(7.51) \quad \Delta_i^* = -r_i^* - \sum_{j=1}^{N-1} p_{ij}^* v_j + g + v_i = 0. \qquad (i \in S)$$

If for all $i \in S$ and $k \in K_i$,

$$(7.52) \quad \Delta_i^k = -r_i^k - \sum_{j=1}^{N-1} p_{ij}^k v_j - g - v_i \geq 0,$$

or using (7.51), for all $i \in S$ and $k \in K_i$,

$$(7.53) \qquad r_i^* + \sum_{j=1}^{N-1} p_{ij}^* v_j \geq r_i^k + \sum_{j=1}^{N-1} p_{ij}^k v_j,$$

we have an optimal solution from the basic theory of linear programming.

While, if there is at least one pair $i \in S$ and $k \in K_i$ such that

$$(7.54) \quad \Delta_i^k = -r_i^k - \sum_{j=1}^{N-1} p_{ij}^k v_j + g + v_i < 0,$$

or

$$(7.55) \qquad r_i^* + \sum_{j=1}^{N-1} p_{ij}^* v_j < r_i^k + \sum_{j=1}^{N-1} p_{ij}^k v_j,$$

there exists an improved solution, or an improved strategy, which corresponds to Policy Improvement Routine. From the viewpoint of linear programming, we may apply $f(i) = k$ such that

$$(7.56) \quad \max_{k \in K_i} \left[r_i^k + \sum_{j=1}^{N-1} p_{ij}^k v_j \right] \; \left(> r_i^* + \sum_{j=1}^{N-1} p_{ij}^* v_j \right)$$

We have already seen from Theorem 7. 7 that substitutions for many variables imply an improved strategy.

Consequently, we have the equivalence between Policy Iteration and Linear Programming Algorithms. That is, Policy Iteratoon Algorithm is a special extension of linear programming such that pivot operations for many (at most N) variables are performed simultaneously.

7. 5. Examples

The examples described in Section 6. 6 are completely ergodic. In this section we shall solve two examples, i.e., the taxicab problem and the automobile replacement problem by using the policy iteration and linear programming algorithms. We further shall propose a new algorithm which is a mexture of the above two algorithms.

Taxicab Problem (Howard [40, p. 44]): The data of the problem were given in Table 6. 2. Starting with an initial strategy we have an optimal strategy with three iterations. The calculations are summarized in Table 7. 2. We also show the linear programming solution starting with the same initial strategy (i.e., the initial basic feasible solution of the policy iteration algorithm. Then we get an optimal solution with 6 steps, where we suppose that 3 steps require to get an initial basic feasible solution.

Now we shall consider a new algorithm which is a mixture of the above two algorithms. We recall the

v_1	1.33	-3.88	-1.18
v_2	7.47	12.85	12.66
v_3	0	0	0
g	9.20	13.15	13.34

V.D.O. P.I.V. V.D.O. P.I.V. V.D.O. P.I.V.

	1	1	2	2	
f	1	2	2	2	stop
	1	2	2	2	

V.D.O.: Value Determination Operation
P.I.V.: Policy Improvement Routine

Table 7. 2. Solution of the taxicab problem by the policy iteration algorithm. (From R. A. Howard, Dynamic Programming and Markov Processes, M. I. T. Press, Cambridge, 1960.)

the relation between the two algorithms. As we have pointed out in Section 6.4, the policy iteration algorithm has a disadvantage to perform its computation. That is, even if there is only one pair $i \in S$ and $k \in K_i$ satisfying (7.55) we must solve the system of N linear equations. While the standard linear programming approach also has a disadvantage, especially, for a large-scale problem. That is, the simplex criteria are always required in each step. But the basic feasible solution may be changed for one variable in each step by using the standard linear programming code. From Policy Improvement Routine we have seen that an improved strategy makes the increase of the average return g. In other words, the pivot operations for many variables yield an improved strategy. Thus we come to a new algorithm. We apply the linear programming approach except that the simplex criteria are used in the same fashion of the policy iteration algorithm. The <u>new algorithm</u> is summarized as follows:

(a) Take an initial strategy and calculate an initial basic feasible solution by using Theorem 7. 7.

(b) Calculate

$$\Delta_i^k = -r_i^k - \sum_{j=1}^{N-1} p_{ij}^k v_j + g + v_i$$

for each $i \in S$ and $k \in K_i$, where we suppose $v_N = 0$.

(c) If $\Delta_i^k \geq 0$ for all $i \in S$ and $k \in K_i$, we get an optima solution. Or, if $\Delta_i^k < 0$ for some $i \in S$ and $k \in K_i$, perform pivot operations (change the basic variables) for each $i \in S$ and its associated one $k \in K_i$ satisfying $\Delta_i^k < 0$ by using the simplex tableau, and return to (b).

This new algorithm can be applied to the discounted Markovian decision model in Chapter 6 by doing the suitable modification. Here we omit the algorithm.

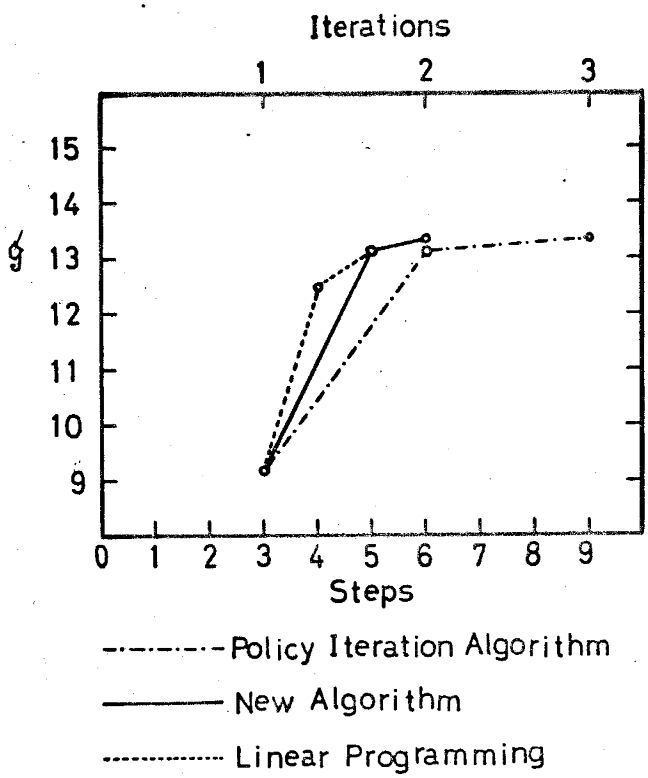

Fig. 7. 3. Comparison among the three algorithms for the taxicab problem.

Fig. 7. 3. shows the results of the threee algorithms, where we assume that 1 iteration corresponds to 3 ($= N$) steps of linear programming. Fig. 7. 3. asserts that the new algorithm is most efficient among the three.

In general the new algorithm is more effecient than the policy iteration one because the calculation of pivot operations such that $\Delta_i^k < 0$ for each $i \in S$ and its associated one $k \in K_i$ is less than that of solving the system of N linear equations. We cannot, however, answer which of the new algorithm or the linear programming algorithm is more efficient.

Automobile Replacement Problem (Howard [40, p. 54]). Let us consider the problem of automobile replacement over a time interval of ten years. We agree to review our current situation every three months and to make a decision on keeping our present car or trading it in at that time. The state of the system, i , is described by the age of the car in three-month periods; i may run from 1 to 40. In order to keep the number of state finite, a car of age 40 remains a car on age 40 forever (it is considered to be essentially worn out). The actions available in each state are these: The first action, $k = 1$, is to keep the present car for another quarter. The other actions, $k > 1$, are to buy a car of age $k - 2$, where $k - 2$ may be as large as 39. We have then 40 states with 41 actions in each state, with the result that there are 41^{40} possible stationary strategies.

The data supplied are the following:

C_i; the cost of buying a car of age i ,

T_i; the trade-in value of a car of age i ,

E_i; the expected cost of operating a car of age until it reaches age $i + 1$,

p_i; the probability that a car of age i will survive to be $i + 1$ without incurring a prohibitively expensive repair.

The probability defined here is necessary to limit the number of states. A car of any age that has a hopeless breakdown is immediately sent to state 40. naturally, $p_{40} = 0$.

The data r_i^k and p_{ij}^k by using the terms of our earlier notation can be written as

$$r_i^k = -E_i \qquad \text{for} \quad k = 1,$$

$$r_i^k = T_i - C_{k-2} - E_{k-2} \qquad \text{for} \quad k > 1,$$

$$p_{ij}^k = \begin{cases} p_i & j = i + 1 \\ 1 - p_i & j = 40 \\ 0 & \text{other} \end{cases} \qquad \text{for} \quad k = 1,$$

$$p_{ij}^k = \begin{cases} p_{k-2} & j = i + 1 \\ 1 - p_{k-2} & j = 40 \\ 0 & \text{other} \end{cases} \qquad \text{for} \quad k > 1.$$

The actual data used in the problem are listed in Table 7. 4 and graphed in Fig. 7. 5. The discontinuities in the cost and trade-in functions were intro duced in order to characterize typical model year

Age in Periods	Cost	Trade-in Value	Operating Expense	Survival Probability	Age in Periods	Cost	Trade-in Value	Operating Expense	Surv Pro bil
i	C_i	T_i	E_i	P_i	i	C_i	T_i	E_i	P
0	$2000	$1600	$50	1.000					
1	1840	1460	53	0.999	21	$345	$240	$115	0.9
2	1680	1340	56	0.998	22	330	225	118	0.9
3	1560	1230	59	0.997	23	315	210	121	0.9
4	1300	1050	62	0.996	24	300	200	125	0.9
5	1220	980	65	0.994	25	290	190	129	0.8
6	1150	910	68	0.991	26	280	180	133	0.8
7	1080	840	71	0.988	27	265	170	137	0.8
8	900	710	75	0.985	28	250	160	141	0.8
9	840	650	78	0.983	29	240	150	145	0.8
10	780	600	81	0.980	30	230	145	150	0.7
11	730	550	84	0.975	31	220	140	155	0.7
12	600	480	87	0.970	32	210	135	160	0.7
13	560	430	90	0.965	33	200	130	167	0.6
14	520	390	93	0.960	34	190	120	175	0.5
15	480	360	96	0.955	35	180	115	182	0.5
16	440	330	100	0.950	36	170	110	190	0.4
17	420	310	103	0.945	37	160	105	205	0.3
18	400	290	106	0.940	38	150	95	220	0.2
19	380	270	109	0.935	39	140	87	235	0.1
20	360	255	112	0.930	40	130	80	250	0

Table 7. 4. Data for the automobile replacement problem. (From R.A. Howard, Dynamic Programming and Markov Processes, M. I. T. Press, Cambridge, 1960.)

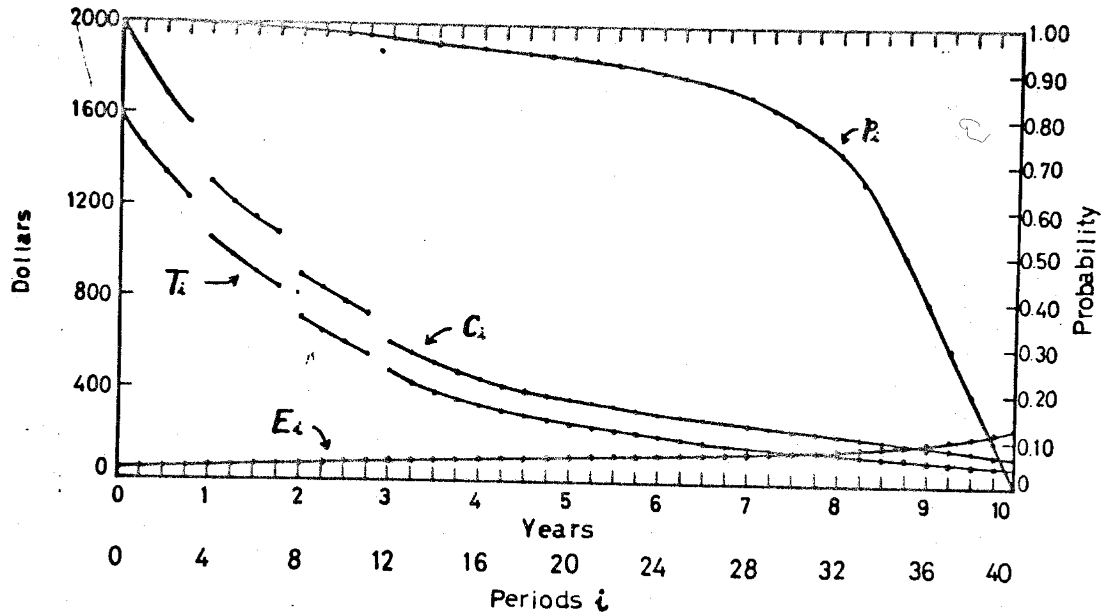

Fig. 7. 5. Graphs for the automobile replacement data. (From R. A. Howard, Dynamic Programming and Markov Processes, M. I. T. Press, Cambridge, 1960.)

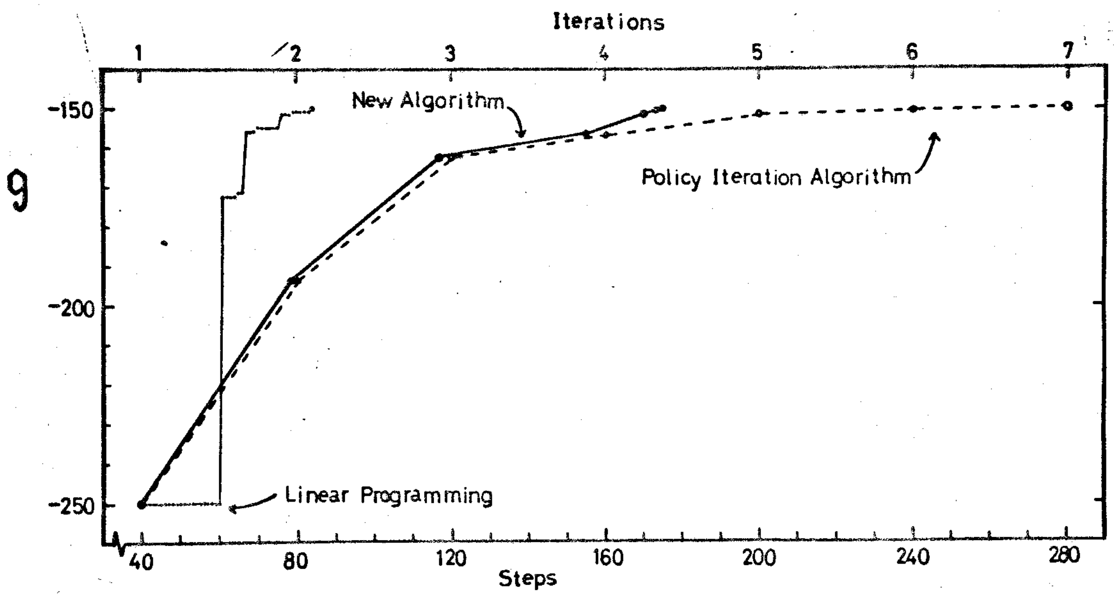

Fig. 7. 6. Comparison among the three algorithms for the automobile replacement problem.

effects.

The policy iteration algorithm yields an optimal strategy with seven iterations. The linear programming algorithm yields an optimal strategy with 84 steps, where we suppose that the calculation of finishing phase I (obtaining an basic feasible solution) needs 40 (= N) steps. The new algorithm yields an optimal strategy with 174 steps. Here we apply the same strategy as an initial one. Fig. 7. 6 shows the results of the above three algorithms.

The new algorithm is more efficient than the policy iteration one. That is, the new algorithm requires the same size criteria, but for the pivot operations we can save the computing time 174/280 (about 62%). Comparing with the linear programming is generally an open question.

7. 6. Terminating Process

In this section we shalll discuss a Markovian decision process of special type. That is, we assume that the system has a common absorbing state whatever decisions we make. Her we denote the absorbing state by state 1. Then we suppose that state 1 is reachable from any other state. This assumption is described precisely as follows:

Terminating Assumption. The common absorbing state is reachable with probability 1 from any state in a finite number of transitions whatever decisions we make.

In other words, this assumption asserts that state 1 is absorbing and state 2, ..., and N are transient whatever decisions we make.

Now our problem is to find a sequence of decisions, namely a strategy of maximizing the total expected return before absorption and its maximum value. This problem is appeared in a Shapley's [55] stochastic game in which the second player is a dummy. This problem is in general a terminating stochastic game. And Shapley's terminating assumption is $\sum_{j=2}^{N} p_{ij}^{k} < 1$ for all i and k. But our assumption is at least $\sum_{j=2}^{N} p_{ij}^{k} < 1$ for some i and its associated all actions $k \in K_i$. Then the system may not move to state 1 in a transition from some transient state. But state 1 is reachable from any transient state in a finite number of transitions with probability 1.

The objective function of our problem is finite because the system is absorbed in a finite number of transitions with probability 1. This is an example in which the total expected return is finite for the nondiscounted model.

The total expected return starting in each state using any strategy π is

$$(7.57) \qquad V(\pi) = \sum_{i=0}^{\infty} P_i(\pi)\, r(f_{i+1}),$$

where $\pi = (f_1, f_2, \ldots, f_i, \ldots)$ and $V(\pi)$, $r(f_i)$ are the $(N - 1) \times 1$ column vectors and $P_i(\pi)$ is the $(N - 1) \times (N - 1)$ matrix elimating the first row and column. For this problem we have the following theorems. The proof is almost similar to that of Section 6. 2. Here we omit the proof.

Lemma 7. 9. $V(\pi^*) \geq V(f, \pi^*)$ for all $f \in F$ implies that π^* is optimal, where (f, π^*) is a strategy preceding π^* with f.

Theorem 7. 10. Exact one of the following has to occur for each $g \in F$.

(i) $V(f^\infty) \geq V(g, f^\infty)$ for all $g \in F$ implies that f^∞ is optimal.

(ii) $V(f^\infty) < V(g, f^\infty)$ for some $g \in F$ implies that $V(f^\infty) < V(g, f^\infty) < V(g^\infty)$.

Theorem 7. 11. There is an optimal strategy which is stationary.

Theorem 7. 10 gives the following policy iteration algorithm:

Value Determination Operation

Take any $f \in F$. Solve

$$v_i = r_i^k + \sum_{j=2}^{N} p_{ij}^k v_j$$

for v_i (i = 2, ..., N), where the superscript k corresponds to the chosen strategy f^∞.

Policy Improvement Routine

Using the values v_i (i = 2, ..., N), find the element of $G(i , f)$ for each i = 2, .., N such that

$$r_i^k + \sum_{j=2}^{N} p_{ij}^k v_j > v_i$$

for all $k \in K_i$. If $G(i , f)$ is empty for all i = 2, .., N, f^∞ is optimal and $V(f^\infty) = [v_i]$ is the total expected return. If at least $g(i) \in G(i , f)$ for some i, make an improved strategy g^∞ such that $g(i) \in G(i , f)$ for some i and $g(i) = f(i)$ for $G(i , f)$ empty, and return to Value Determination Operation.

We have also the following linear programming problem for the terminating process:

$$\text{Max} \sum_{j=2}^{N} \sum_{k \in K_j} r_j^k x_j^k$$

subject to

$$(7.59) \quad \sum_{k \in K_j} x_j^k - \sum_{i=2}^{N} \sum_{k \in K_i} p_{ij}^k x_i^k = a_j , \qquad (j = 2, \ldots, N)$$

$$(7.60) \quad x_j^k \geq 0. \qquad (j = 2, \ldots, N; \ k \in K_j)$$

where $a = (a_1 , a_2, \ldots, a_N)$ is an initial distribution of the process.

This derivation of the above linear programming problem is almost similar to that of Section 6. 3. Thus the same properties discussed in Section 6. 3 are also satisfied. Further we can show the relation between the above two algorithms. These facts can be easily verified. We don't state here in detail.

7. 7. Policy Iteration Algorithm for the General Case

In the preceding sections we have discussed the special structure processes, i.e., the completely ergodic process and the terminating process. In this and the next sections we shall consider the general process that there may be several ergodic sets and some transient states, and these sets may vary from strategy to strategy.

In this section we shall derive the policy iteration algorithm for the general process. Here we use the results of the processes with discounting in the preceding chapter.

We shall treat $\beta = 1$ as a limiting case of $\beta < 1$. For the discounted model, the discounted total expected return is $V_\beta(\pi) = \sum_{i=0}^{\infty} \beta^i P_i(\pi)\, r(f_{i+1})$, where $0 \leq \beta < 1$. Let $\pi(\beta)$ be β-optimal and $U(\beta) = V_\beta(\pi(\beta))$.

Definition 7. 12. A strategy π^* is 1-optimal if

(7.61) $$\lim_{\beta \to 1-0} [V_\beta(\pi^*) - U(\beta)] = 0.$$

It will be shown that 1-optimal strategies are important for our problem. Now we shall show the following theorem, which corresponds to Lemma 7. 6. The proof is found in Blackwell [9].

Theorem 7. 13. For any stationary strategy f^∞, let $P^*(f) = \lim_{n\to\infty} \sum_{i=0}^{n-1} P^i(f)/n$, which has been defined in Lemma 7. 3. Then

$$(7.62) \qquad V_\beta(f^\infty) = \frac{u(f)}{1-\beta} + v(f) + \epsilon(\beta, f),$$

where $u(f)$ is a unique solution of

$$(7.63) \qquad (I - P(f))u = 0, \quad P^*(f)u = P^*(f)r(f),$$

$v(f)$ is a unique soltuion of

$$(7.64) \qquad (I - P(f))v = r(f) - u(f), \quad P^*(f)v = 0,$$

and $\epsilon(\beta, f) \to 0$ as $\beta \to 1 - 0$.

The solutions $u(f)$ and $v(f)$ of (7.63) and (7.64) may be written by

$$(7.65) \qquad u(f) = P^*(f)r(f),$$

$$(7.66) \qquad v(f) = H(f)r(f) = [(I - P(f) + P^*(f))^{-1} - P^*(f)]r(f).$$

The above equations state that, specifying $f \in F$, we can determine $P^*(f)$ and also obtain $u(f)$ and $v(f)$. Thus our problem is to find a strategy which maximizes $u(f)$ (and also $v(f)$) within all stationary strategies. That is, our problem is a combinatorial one. We then shall need the efficient algorithm for this problem.

As we have seen in Theorem 7. 4. that there exists an optimal stationary strategy under the average return criterion, we define the set of strategies such that

(7.67) $$F' = \{ f \mid f \in F, u(f) \geq u(g) \text{ for all } g \in F \}.$$

That is, F' is the set of all $f \in F$ having maximal average return per unit time. Further, we define the set of strategies such that

(7.68) $$F'' = \{ f \mid f \in F', v(f) \geq v(g) \text{ for all } g \in F' \}.$$

It is evident that F'' is a subset of F' and is the set of all $f \in F'$ having maximal bias term $v(f)$.

For the discounted model, $V_\beta(g, f^\infty)$ is computed by using (7.62) as follows:

(7.69) $$\begin{aligned} V_\beta(g, f^\infty) &= r(g) + \beta P(g) V_\beta(f^\infty) \\ &= \frac{\beta P(g) u(f)}{1-\beta} + r(g) + \beta P(g) v(f) + \beta P(g) \varepsilon(\beta, f) \\ &= \frac{P(g) u(f)}{1-\beta} + r(g) - P(g) u(f) + P(g) v(f) \\ &\qquad - (1-\beta) P(g) v(f) + \beta P(g) \varepsilon(\beta, f) \\ &= \frac{P(g) u(f)}{1-\beta} + r(g) - P(g) u(f) + P(g) v(f) + \varepsilon_1(\beta, f, g), \end{aligned}$$

where $\varepsilon_1(\beta, f, g) = -(1-\beta) P(g) v(f) + \beta P(g) \varepsilon(\beta, f) \to 0$ as $\to 1-0$.

Comparing the right-hand side of both (7.62) and (7.63), we define the set of actions such that

(7.70) $\mathcal{G}(i,f)=\{k \mid k\in K_i,\ \sum_{j\in S} p_{ij}^{k} u_j > u_i,\ \text{or} \sum_{j\in S} p_{ij}^{k} u_j = u_i$

$\text{and } r_i^{k}+\sum_{j\in S} p_{ij}^{k} v_j > u_i + v_i\},$

where u_i and v_i are the ith elements of $u(f)$ and $v(f)$, respectively. Further we define

(7.71) $$\mathcal{G}(f)=\mathop{\times}_{i=1}^{N} \mathcal{G}(i,f).$$

Using the above results and notations, we have the following theorem which has been given by Blackwell [9].

Theorem 7. 14. Take any $f\in F$.

(a) If $\mathcal{G}(f)$ is empty, then $f\in F'$.

(b) If $g(i)\in \mathcal{G}(i,f)$ for some $i\in S$ and $g(i) = f(i)$ whenever $g(i)\notin \mathcal{G}(i,f)$, then $V_\beta(g^\infty) > V_\beta(f^\infty)$ for all β (<1) sufficiently near 1.

This theorem describes the Policy Iteration Algorithm, which yields an optimal stationary strategy $f\in F'$ with a finite number of iterations.

7. 8. Linear Programming Considerations on the General Case

In this section we shall develop a linear programming algorithm for the general Markovian decision process with no discounting. In this section we apply an approach of treating directly the long-run average return per unit time. From Theorem 7. 4, we restrict our attention to stationary strategies and wirte a stationary strategy f instead of f^{∞}.

From (7.10) we have for any f

$$(7.72) \qquad a\Gamma(f) = aP^{*}(f)r(f) = \sum_{i\in S}\sum_{j\in S} a_i p_{ij}^{*}(f) r_j(f),$$

which is the average return per unit time starting in an initial distribution a, where $P^{*}(f) = [p_{ij}^{*}(f)]$. Note that an optimal strategy under the average return criterion is independent of the initial distribution since an optimal strategy is attained simultaneously for each initial state. So, we consider a problem for finding a stationary strategy which maximizes (7.72) under all $f \in F$. That is, our problem is to find an optimal strategy f such that

$$(7.73) \qquad \max_{f\in F} \sum_{i\in S}\sum_{j\in S} a_i p_{ij}^{*}(f) r_j(f).$$

It is convenient to extend of decisions to include randomized strategies. So, let d_j^k denote the joint probability that the system is in state $j \in S$ and an action $k \in K_j$ is made. It is evident that

$$(7.74) \qquad d_j^k \geqslant 0 \quad (j \in S, k \in K_j), \qquad \sum_{k\in K_j} d_j^k = 1.$$

We consider any fixed nonrandomized strategy for the latter discussion.

Setting

(7.75) $$x_j^k = \sum_{i\in S} a_i p_{ij}^*(f)\, d_j^k \qquad (j\in S, k\in K_j),$$

and

(7.76) $$y_j^k = \sum_{i\in S} a_i h_{ij}(f)\, d_j^k \qquad (j\in S, k\in K_j),$$

where

(7.77) $$[h_{ij}(f)] = H(f) = (I - P(f) + P^*(f))^{-1} - P^*(f),$$

and using the relations $P^*(I - P) = 0$ (from (7.3)) and $P^* + H(I - P) = I$ (from (7.6)), we have

(7.78) $$\sum_{j\in S}\sum_{k\in K_j} (\delta_{jl} - p_{jl}^k)\, x_j^k = 0 \qquad (l\in S),$$

and

(7.79) $$\sum_{k\in K_l} x_l^k + \sum_{j\in S}\sum_{k\in K_j} (\delta_{jl} - p_{jl}^k)\, y_j^k = a_l \qquad (l\in S),$$

where δ_{jl} is the Kronecker's delta. It is evident from (6.3), (7.74), and (7.77) that

(7.80) $$x_j^k = \sum_{i\in S} a_i q_{ij}^*(f)\, d_j^k \geqslant 0 \qquad (j\in S, k\in K_j).$$

While the sign of y_j^k is not clear and we shall show afterwards that $y_j^k \geq 0$ for any transient state.

Thus we have a following linear programming problem:

(7.81) Max $\sum_{j\in S}\sum_{k\in K_j} r_j^k x_j^k$

subject to

(7.82) $$\sum_{j\in S}\sum_{k\in K_j}(\delta_{jl}-p_{jl}^k)\,x_j^k=0 \qquad (l\in S),$$

(7.83) $$x_j^k\geqslant 0 \qquad (j\in S,\,k\in K_j),$$

(7.84) $$\sum_{k\in K_l} x_l^k+\sum_{j\in S}\sum_{k\in K_j}(\delta_{jl}-p_{jl}^k)\,y_j^k=a_l \qquad (l\in S).$$

We shall afterwards show that $y_j^k \geq 0$ for any transient state j.

For a fixed strategy f, Markov matrix $P(f)$ has some ergodic sets plus a transient set. Appropriately relabeling the number of states, we have the following form for Markov Matrix $P(f)$:

(7.85) $$P(f)=\left(\begin{array}{cccc|c} P_{11} & & & 0 & \\ & P_{22} & & & 0 \\ & & \ddots & & \\ 0 & & & P_{\nu\nu} & \\ \hline P_{\nu+1,1} & P_{\nu+1,2} & \cdots & P_{\nu+1,\nu} & P_{\nu+1,\nu+1} \end{array}\right),$$

where P_{11}, ..., $P_{\nu\nu}$ are submatrices associated with each <u>ergodic set</u> E_μ (μ = 1, .., ν), respectively, and the remaining states specify a set T of transient states.

Next two lemmas are useful to eliminate redundant constraints in (7.82) and (7.84). In the discussion of these lemmas (containing Lemma 7.17) we have to restrict ourselves to any fixed nonrandomized strategy since we cannot consider simultaneouly state classification for randomized strategies. Thus we omit the summation on k below.

Lemma 7. 15. Take any $f \in F$. For any $j \in E_\mu$ ($\mu = 1, \ldots, \nu$), constraints (7.84) become

$$(7.86) \qquad \sum_{l \in E_\mu} x_l^k + \sum_{l \in E_\mu} \sum_{j \in T} (\delta_{jl} - p_{jl}^k) y_j^k = \sum_{l \in E_\mu} a_l \qquad (\mu = 1, 2, \ldots, \nu).$$

Proof. Using Lemma 7. 3 (c), we combine constraints (7.84) by summing on E_μ. While, $\sum_{i \in S} = \sum_{i \in E_\mu} + \sum_{\substack{\nu=1 \\ \nu \neq \mu}}^{\nu} \sum_{i \in E_\nu} + \sum_{i \in T}$ and $\sum_{i \in E_\mu} (\delta_{ij} - p_{ij}^k) = \sum_{\substack{i \in E_\nu \\ \nu \neq \mu}} (\delta_{ij} - p_{ij}^k) = 0$ (from (7.85)) implies (7.86), which completes the proof.

Lemma 7. 16. For any ergodic set E_μ ($\mu = 1, \ldots, \nu$), one of constraints (7.82) associated with E_μ is redundant.

Proof. It is obvious from the property of Markov chains and the fact that $P_{\mu\mu}$ ($\mu = 1, \ldots, \nu$) is a Markov matrix, which completes the proof.

The above two lemmas give the necessary constraints for any ergodic set. Next lemma also gives a constraint for any transient state.

Lemma 7. 17. Take any $f \in F$. For any state $j \in T$, a constraint (7.84) becomes

(7.87) $$\sum_{j \in T} (\delta_{jl} - p_{jl}^k) y_j^k = a_l \qquad (l \in T),$$

where

(7.88) $$y_l^k \geqslant 0 \qquad (l \in T, k = f(l)).$$

Proof. From $[p_{ij}^*(f)]_{i \in S, j \in T} = [0]_{i \in S, j \in T}$ and (7.75), we have $x_l^k = 0$ for any $l \in T$ and $k = f(l)$. And we have

$$\sum_{i \in S} (\delta_{ij} - p_{ij}^k) y_i^k = \sum_{i \notin T} (\delta_{ij} - p_{ij}^k) y_i^k + \sum_{i \in T} (\delta_{ij} - p_{ij}^k) y_i^k$$

$= \sum_{i \in T} (\delta_{ij} - p_{ij}^k) y_i^k = a_j$, which corresponds to (7.87).

While, from $[h_{ij}(f)]_{i,j \in T} = [I - P_{\nu+1,\nu+1}]^{-1} = \sum_{n=0}^{\infty} P_{\nu+1,\nu+1}^n \geq 0$, $[h_{ij}(f)]_{i \notin T, j \in T} = [\sum_{n=0}^{\infty} (P^n(f) - P^*(f)]_{i \notin T, j \in T} = [0]_{i \notin T, j \in T}$ and (7.76), we have

(7.89) $$y_l^k = \sum_{i \in S} a_i h_{il}(f) d_l^k \geqslant 0 \qquad (l \in T, k = f(l)),$$

which completes the proof.

From (7.88), we suppose that $y_j^k \geq 0$ for any $j \in S$, $k \in K_j$, since y_j^k disappears for any ergodic state j.

Thus we also consider a dual problem of the linear programmming problem (7.81), (7.82), (7.83), (7.84), and (7.88). Let the $N \times 1$ column vectors $u(f)$ and $v(f)$ be the corresponding dual variables. Then its dual problem is:

(7.90) Max $\sum_{i\in S} a_i u_i(f)$

subject to

(7.91) $$u_i(f) \geqslant \sum_{j\in S} p_{ij}^k u_j(f) \qquad (i \in S,\ k \in K_i),$$

(7.92) $$u_i(f) + v_i(f) \geqslant r_i^k + \sum_{j\in S} p_{ij}^k v_j(f) \qquad (i \in S,\ k \in K_i),$$

(7.93) $u_i(f),\ v_i(f)$; unconstrained in sign ($i \in S$),

where u_i, v_i are the ith elements of $u(f)$, $v(f)$, respectively. We know that this dual problem corresponds to the policy iteration algorithm, i.e, this dual problem is immediately derived from the policy iteration algorithm. We also note that the dual variables $u(f)$ and $v(f)$ are unique solutions of

(7.94) $$u(f) = P(f)\,u(f), \qquad u(f) + v(f) = r(f) + P(f)\,v(f),$$

where from (7.7) we set the value of one v_i in each ergodic set to zero, which refers to (7.86). Then $v(f)$ is a <u>relative solution</u> and the difference be tween the exact solution of (7.64) and $v(f)$ in (7.94) is a constant.

Now we have a linear programming algorithm for the general Markovian decision processes. The algorithm is made by using Lemmas 7. 3, 7. 15, 7. 16, and 7. 17. Further we note that the dual variables ($u(f)$, $v(f)$) are also simplex multipliers. Using these simplex multipliers, we have the simplex criterion, which corresponds to Policy Improvement Routine in the policy iteration algorithm. The direct proof of increasing the average return by its simplex criterion without linear programming properties can be made. It is convenient to consider the following set of actions which corresponds to the simplex criterion of the primal problem using simplex multipliers ($u(f)$, $v(f)$):

$$G(i,f) = \Big\{ k \in K_i \mid \sum_{j\in S} p_{ij}^k u_j(f) > u_i(f), \text{ or } \sum_{j\in S} p_{ij}^k u_j(f) = u_i(f) \text{ and } r_i^k + \sum_{j\in S} p_{ij}^k v_j(f) > u_i(f) + v_i(f) \Big\}. \tag{7.95}$$

Then we have the following proposition of describing the linear programming algorithm without proof.

Proposition 7. 18. Taking any $f \in F$ and determining the constraints for each state according to the state classification (using Lemmas 7. 15, 7. 16, and 7. 17), we can obtain a basic feasible solution which corresponds to a strategy f and gives its dual variables $u(f)$ and $v(f)$ (simplex multipliers). Using these simplex multipliers, we have the simplex criterion. That is, $G(i, f)$ is empty for all $i \in S$, then we have an optimal stationary strategy. Otherwise, select a new strategy g such that $g(i) \in G(i, f)$ and $g(i) =$

$f(i) \notin G(i, f)$. Returning to the first part of this proposition, repeat until an optimal strategy is obtained.

Note that Proposition 7. 18 describes a special structure linear programming algorithm such that pivot operations for many variables are performed simultaneously if there are two or more states such that $G(i, f)$ is nonempty.

Proposition 7. 18 and primal and dual problems imply the following corollary.

Corollary 7. 19. The policy iteration algorithm is equivalent to the linear programmong one.

Proof. Finding a basic feasible solution (i.e., the dual variables) corresponds to Value Determination Operation, and the simplex criterion of the next step corresponds to Policy Improvement Routine. But in the policy iteration algorithm pivot operations are performs simultaneously for many variables. The facts complete the proof.

Next corollary is clear if we consider the policy iteration algorithm.

Corollary 7. 20. An optimal strategy is independent of the initial distribution a.

Here we shall show the linear programming algorithm and the policy iteration algorithm for Howard's example [40, p. 65]. The data of the problem are given in Table 7. 7. Let an initial strategy denote by f , its associated Markov matrix by $P(f)$ and the return vector by $r(f)$. Then

$$f=\begin{bmatrix}3\\2\\1\end{bmatrix},\quad P(f)=\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix},\quad r(f)=\begin{bmatrix}3\\4\\8\end{bmatrix}.$$

Primal and dual problems are written by the following tableau (reduced Tucker Diagram), where only the data associated with a basic feasible solution are given. And the superscripts of its chosen strategy are omitted.

$$\begin{array}{r|ccc|ll}
 & x_1 & x_2 & x_3 & & \\
\hline
v_1 & 1 & 0 & -1 & =0 & \\
u_2 & 0 & 1 & 0 & =a_2 & (v_2=0) \\
u_1=u_3 & 1 & 0 & 1 & =a_1+a_3 & (v_3=0) \\
\hline
 & \leqq & \leqq & \leqq & & \\
 & 3 & 4 & 8 & &
\end{array}$$

Dual Variables

$$u(f)=\begin{bmatrix}11/2\\4\\11/2\end{bmatrix},\quad v(f)=\begin{bmatrix}-5/2\\0\\0\end{bmatrix}.$$

Using the simplex criterion (or equivalently Policy Improvement Routine), we have a next improved strategy g and its data.

State	Action	Probability			Return
i	k	p_{i1}^k	p_{i2}^k	p_{i3}^k	r_i^k
1	1	1	0	0	1
	2	0	1	0	2
	3	0	0	1	3
2	1	1	0	0	6
	2	0	1	0	4
	3	0	0	1	5
3	1	1	0	0	8
	2	0	1	0	9
	3	0	0	1	7

Table 7. 7. Data for a multichain example. (From R. A. Howard, Dynamic Programming and Markov Processes, M. I. T. Press, Cambridge, 1960.)

$$g = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}, \quad P(g) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad r(g) = \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}.$$

Then

$$\begin{array}{r|ccc|ll} & y_1 & y_2 & x_3 & & \\ \hline v_1 & 1 & 0 & 0 & = a_1 & \\ v_2 & 0 & 1 & 0 & = a_2 & \\ u_1 = u_2 = u_3 & 1 & 1 & 1 & = a_3 & \quad (v_3 = 0) \\ \hline & \leq & \leq & \leq & & \\ & 3 & 5 & 7 & & \end{array}$$

Dual Variables

$$u(g) = \begin{bmatrix} 7 \\ 7 \\ 7 \end{bmatrix}, \quad v(g) = \begin{bmatrix} -4 \\ -2 \\ 0 \end{bmatrix}.$$

Using the simplex criterion, we have an optimal strategy g since $G(i, g)$ is empty for any $i \in S$.

CHAPTER VIII

MARKOV RENEWAL PROGRAMMING

8. 1. Introduction

In the preceding chapters we have discussed the decision processes in which the decisions are made synchronously in each time period. In this chapter we shall discuss the decision processes in which the sojourn time in each state is random and thus the decisions are made asynchronously. That is, our preceding models are based on a discrete time Markov chain. But in this chapter we shall consider a decision process based on a continuous time stochastic process.

An extension of a discrete time Markov chain is a continuous time Markov process. A decision process based on a continuous time Markov process has been studied by Howard [40]. But in this chapter we shall treat a decision process based on a Markov renewal process which includes a continuous time Markov process as a special case.

A Markov renewal process of a semi-Markov process is a marriage of Markov processes and renewal processes. A Markov renewal process, roughly speaking, is a stochastic process which moves from one state to another with given probability laws, but the sojourn time in a state is a random variable with distribution depending on that state and the next visiting state. Markov renewal processes include a renewal process, a discrete time Markov chain and a continuous time Markov process as special cases. That is, a renewal process is a Markov renewal process with one state. A discrete time Markov chain is a Markov renewal process in which every sojourn time distribution is degenerate at unit time, and a continuous time Markov process is a Markov renewal process in which all distributions are exponential.

In this chapter we shall treat a decision process based on a Markov renewal process or a semi-Markov process, i.e., a Markov renewal program, or a semi-Markovian decision process.

In Section 8. 2 we shall study the properties of Markov renewal processes, and further consider Markov renewal processes with returns in Section 8. 3. In the successive sections we shall introduce the decisions of Markov renewal processes with returns and show the similar results that Markov renewal programs are formulated by linear programming problems for discounted and nondiscounted models. In the discussion we shall derive the corresponding policy iteration algorithms from these linear programming problems.

8. 2. Markov Renewal Processes

In this section we shall show the known facts about Markov renewal processes or semi-Markov processes which will be useful or suggestive for the latter discussion. Semi-Markov processes were first discussed by Lévy, Smith, and Takács, independently, in 1954.

Consider a system that moves from state to state. We define a stochastic process $\{ Z_t; \ t \geq 0 \}$, where

$Z_t = i$ denotes that the system is in state i at time t. The states are denoted by integers $i = 1, 2, \ldots, N \in S$ and in general the state space S is countable, but we consider only the case that S is finite throughout this chapter. In the discussion below a part of results is valid for the countable case, but we shall state nothing for these facts. Note that we shall introduce the concept of regularity for each state afterwards.

The probability laws of state transitions obey an $N \times N$ Markov matrix $P = [\, p_{ij} \,]$ which is called an imbedded Markov chain, where p_{ij} is the probability from state i to state j. Further, the sojourn time in any state i is a random variable with distribution $F_{ij}(t)$ which depends both on that state and on the next visiting state j. We define

(8.1) $$Q_{ij}(t) = p_{ij} F_{ij}(t), \qquad (i, j \in S)$$

where $Q_{ij}(t)$ satisfies

(8.2) (i) $$Q_{ij}(0) = 0, \qquad (i, j \in S)$$

(8.3) (ii) $$\sum_{j \in S} Q_{ij}(\infty) = \sum_{j \in S} p_{ij} = 1. \qquad (i \in S)$$

We define an $N \times N$ matrix Q with element $Q_{ij}(t)$, which is called a matrix of transition distributions. Further, we define

(8.4) $$H_i(t) = \sum_{j \in S} Q_{ij}(t), \qquad (i \in S)$$

which is a distribution of the sojourn time in state i disregarding next designating state. So, it is called a distribution of unconditional sojourn time, or an unconditional distribution. Let a be an initial distribution ($1 \times N$ row vector);

$$(8.5) \qquad a = (a_1, a_2, \ldots, a_N),$$

where

$$(8.6) \qquad \sum_{i \in S} a_i = 1, \quad a_i \geq 0. \qquad\qquad (i \in S)$$

Then, the stochastic process $\{Z_t ; t \geq 0\}$ is called a semi-Markov process determined by

$$(8.7) \qquad (N, a, Q).$$

While, we are concerned with an N-dimensional renewal quantity

$$(8.8) \qquad \mathcal{N}(t) = (\mathcal{N}_1(t), \mathcal{N}_2(t), \ldots, \mathcal{N}_N(t)),$$

then the stochastic process $\{\mathcal{N}(t);\ t \geq 0\}$ is called a Markov renewal process. A semi-Markov process is regular if each state is entered only a finite number of times within a finite time with probability 1, i.e., $\mathrm{Prob}[\mathcal{N}_i(t) < \infty] = 1$ for each $i \in S$ and $t \geq 0$. Thus a regular semi-Markov process has always a finite number of transitions within a finite time. We consider only a regular semi-Markov process throughout this chapter. We suppose that each state is defined such that the semi-Markov process is regular.

We define

(8.9) $$P_{ij}(t) = P_r[Z_t = j \mid Z_0 = i], \qquad (i, j \in S)$$

which denotes the probability that the system is in state j at time t given that the system is in state i at time zero. Further we define

(8.10) $$G_{ij}(t) = P_r[\mathcal{N}_j(t) > 0 \mid Z_0 = i], \qquad (i, j \in S)$$

(8.11) $$M_{ij}(t) = E[\mathcal{N}_j(t) \mid Z_0 = i], \qquad (i, j \in S)$$

where $E[\cdot]$ denotes the expectation. Here $G_{ij}(t)$ is the first passage time distribution from state i to state j, and $M_{ij}(t)$ is the mean number of visits to state j up to time t starting in state i, which refers to a renewal function in the renewal theory.

From renewal-theoretic considerations we have immediately

(8.12) $$\begin{aligned} P_{ij}(t) &= (1 - H_i(t))\,\delta_{ij} + \sum_{k \in S} Q_{ik}(t) * P_{kj}(t) \\ &= (1 - H_i(t))\,\delta_{ij} + G_{ij}(t) * P_{jj}(t), \end{aligned}$$

(8.13) $$G_{ij}(t) = Q_{ij}(t) + \sum_{\substack{k \in S \\ k \neq j}} Q_{ik}(t) * G_{kj}(t),$$

(8.14) $$\begin{aligned} M_{ij}(t) &= G_{ij}(t) + G_{ij}(t) * M_{jj}(t) \\ &= Q_{ij}(t) + \sum_{k \in S} Q_{ik}(t) * M_{kj}(t), \end{aligned}$$

where $*$ denotes the convolution. To solve the above equations with respect to $P_{ij}(t)$, $G_{ij}(t)$ and $M_{ij}(t)$, we apply the Laplace-Stieltjes transforms. Let the small letter with argument s denote the Laplace-Stieltjes transform of the corresponding quantity, e.g., $p_{ij}(s)$ is the Laplace-Stieltjes transform of $P_{ij}(t)$. Further, let $g(s)$ and $m(s)$ be $N \times N$ matrices with elements $g_{ij}(s)$ and $m_{ij}(s)$, respectively.

Taking the Laplace-Stieltjes transforms for (8.12), (8.13) and (8.14), and solving to each quantity, we have

$$m(s) = [I - g(s)]^{-1} g(s) = [I - g(s)]^{-1} - I, \tag{8.15}$$

$$g_{ij}(s) = m_{ij}(s) / [1 + m_{jj}(t)], \tag{8.16}$$

$$p_{ij}(s) = p_{jj}(s)\, g_{ij}(s), \qquad (i \neq j) \tag{8.17}$$

$$p_{jj}(s) = [1 - h_j(s)] / [1 - g_{jj}(s)]. \tag{8.18}$$

Thus, inverting the above equations successively, we have the required quantities, $M_{ij}(t)$, $G_{ij}(t)$ and $P_{ij}(t)$.

Then we can give the state classification for a semi-Markov process. We first define that two states i and j are said to communicate if $G_{ij}(\infty)\, G_{ji}(\infty) > 0$. Since the communication relation is an equivalent one, we may apply the usual state classification for finite Markov chains, which obeys that of Kemeny-Snell [47].

Let

$$b_{ij} = \int_0^\infty t\, dF_{ij}(t), \qquad (i, j \in S) \tag{8.19}$$

and

$$\eta_i = \int_0^\infty t\, dH_i(t) = \sum_{j \in S} p_{ij} \int_0^\infty t\, dF_{ij}(t) = \sum_{j \in S} p_{ij} b_{ij}, \quad (i \in S) \tag{8.20}$$

be mean values of $F_{ij}(t)$ and $H_i(t)$, respectively. All these mean values are assumed to be finite. Further we shall give the additioanl assumptions if necessary.

Let μ_{ij} and $\mu_{ij}^{(2)}$ be the first and second moments of $G_{ij}(t)$. From the mean-theoretic considerations, we have

$$\mu_{ij} = \sum_{\substack{k \in S \\ k \neq j}} p_{ik}\mu_{kj} + \eta_i, \tag{8.21}$$

$$\mu_{ij}^{(2)} = \sum_{\substack{k \in S \\ k \neq j}} p_{ik} [\mu_{kj}^{(2)} + 2 b_{ik}\mu_{kj}] + \eta_i^{(2)}, \tag{8.22}$$

where

$$\eta_i^{(2)} = \int_0^\infty t^2 dH_i(t). \qquad (i \in S) \tag{8.23}$$

If the imbedded Markov chain is ergodic and let π_j be the limiting probability of the imbedded Markov chain, we have

$$\mu_{jj} = \sum_{k \in S} \pi_k \eta_k / \pi_j, \tag{8.24}$$

(8.25) $\mu_{jj}^{(2)} = [\sum_{k\in S} \pi_k \eta_k^{(2)} + 2 \sum_{\substack{k\in S\\ k\neq j}} \pi_i b_{ik} \mu_{kj}] + \pi_j,$

where μ_{jj} and $\mu_{jj}^{(2)}$ are the first and second moments of the reccurrence time in state j .

We consider the limiting (or steady state) probability $P_{ij}(\infty)$. In general, we have

(8.26) $P_{ij}(\infty) = G_{ij}(\infty)\,\eta_j / \mu_{jj}\,.$

In particular, if the semi-Markov process is ergodic, from $G_{ij}(\infty) = 1$ for all $i, j \in S$ and (8.24), we have

(8.27) $P_{ij}(\infty) = \eta_j / \mu_{jj} = \pi_j \eta_j \Big/ \sum_{k\in S} \pi_k \eta_k\,,$

which is independent of any initial distribution. In general we can obtain $G_{ij}(\infty)$ and μ_{jj} for any $i, j \in S$, and thus we can obtain the limiting probability $P_{ij}(\infty)$ for any $i, j \in S$. Here we omit the results.

Let n_i be the mean time to absorption, starting in any transient state $i \in T$. Here the word "absorption" means the entrance of some ergodic state. We have

(8.28) $n_i = \sum_{j\notin T} p_{ij} b_{ij} + \sum_{j\in T} p_{ij} (b_{ij} + n_j)$

$= \eta_i + \sum_{j\in T} p_{ij} n_j\,.$

Let n and η be column vectors with n_i and η_i , respectively, where $i \in T$. Then we have

$$n = \eta + P_T n, \tag{8.29}$$

or

$$n = [I - P_T]^{-1} \eta, \tag{8.30}$$

where P_T is a submatrix of P eliminating all ergodic states.

Finally we shall show the asymptotic behavior about the generalized renewal quantities $M_{ij}(t)$. From (8.16), we have

$$\begin{aligned} m_{ij}(s) &= g_{ij}(s)\,[1 + m_{jj}(s)] \\ &= g_{ij}(s)\left[1 + \frac{g_{jj}(s)}{1 - g_{jj}(s)}\right]. \end{aligned} \tag{8.31}$$

If states i and j are in the same ergodic set, expanding $g_{ij}(s)$ at a small s, we have, for a small s,

$$m_{ij}(s) = \frac{1}{s\,\mu_{jj}} + \frac{\mu_{ij}^{(2)}}{2\,(\mu_{jj})^2} - \frac{\mu_{ij}}{\mu_{jj}} + O(1). \tag{8.32}$$

Thus from a Tauberian theorem we have

$$M_{ij}(t) - \frac{t}{\mu_{jj}} \longrightarrow \frac{\mu_{jj}^{(2)}}{2\,(\mu_{jj})^2} - \frac{\mu_{ij}}{\mu_{jj}}\ , \tag{8.33}$$

which is an analogous result of an extension of the so-called elementary renewal theorem. We note that, if the semi-Markov process considered is periodic, the limit of (8.26), (8.27), (8.30) and (8.31) means Césaro limit, conversely if it is nonperiodic, the limit is the usual one.

8. 3. Markov Renewal Process with Returns

In this section we shall consider the returns associated with Markov renewal processes. In the discrete time case we have considered the return per time period, but in the continuous time case we may consider the returns r_i ($i \in S$). That is, when the system is in state i during a unit time, we receive the return r_i. Thus when the system is in state i during a time duration t, we receive the return $r_i t$.

First, we shall consider the discounted case. Since the process is continuous in time, we use a discount factor α ($\alpha > 0$) of exponential type. That is, if we have a unit return at any time, we have the return $e^{-\alpha t}$ after a time duration t from that time. Then, if we have the return r_i, the accumulated return between 0 and t is

$$(8.34) \qquad \int_0^t r_i e^{-\alpha\tau} d\tau = \frac{r_i}{\alpha}[1 - e^{-\alpha t}].$$

Let $v_i(t)$ be the discounted total expected return up to time t starting in state i at time zero. Then we have similarly the following equation of renewal type with respect to $v_i(t)$ (e.g., confer (8.12)):

$$(8.35)\quad v_i(t)=[1-H_i(t)]\frac{r_i}{\alpha}[1-e^{-\alpha t}] + \sum_{j\in S}\int_0^t\left\{\frac{r_i}{\alpha}[1-e^{-\alpha\tau}]+e^{-\alpha\tau}v_j(t-\tau)\right\}dQ_{ij}(\tau). \qquad (i\in S)$$

Since the discounted total expected return $v_i(t)$ converges as $t\to\infty$ to a finite value, we set

$$(8.36)\qquad v_i(\alpha)=\lim_{t\to\infty} v_i(t), \qquad (i\in S)$$

(note that $v_i(\alpha)$ is a function of α) and letting $t\to\infty$ in (8.35), we have

$$(8.37)\quad v_i(\alpha)=\frac{r_i}{\alpha}[1-h_i(\alpha)]+\sum_{j\in S} q_{ij}(s)v_j(\alpha), \qquad (i\in S)$$

where $h_i(\alpha)$ and $q_{ij}(s)$ are the Laplace-Stieltjes transforms of $H_i(t)$ and $Q_{ij}(t)$ evaluated at $s=\alpha$, respectively. Setting

$$(8.38)\qquad \rho_i(\alpha)=\frac{r_i}{\alpha}[1-h_i(\alpha)], \qquad (i\in S)$$

and defining two $N\times 1$ column vectors

$$(8.39)\qquad \rho(\alpha)=\begin{bmatrix}\rho_1(\alpha)\\ \vdots\\ \rho_N(\alpha)\end{bmatrix}, \qquad v(\alpha)=\begin{bmatrix}v_1(\alpha)\\ \vdots\\ v_N(\alpha)\end{bmatrix},$$

we have

$$(8.40) \qquad v(\alpha) = \rho(\alpha) + q(\alpha)v(\alpha),$$

or

$$(8.41) \qquad v(\alpha) = [I - q(\alpha)]^{-1}\rho(\alpha),$$

since we have known that $I - q(\alpha)$ is nonsingular for $\alpha > 0$. Thus we have the discounted total expected return starting in each state $i \in S$. If the system starts in an initial distribution a in (8.5), we have

$$(8.42) \qquad a v(\alpha) = a\,[I - q(\alpha)]^{-1}\rho(\alpha).$$

Second, we shall consider the nondiscounted case. We have seen two approaches of the nondiscounted model in the discrete time case. For the continuous model we have also two approaches. One is treating as a limiting case of $\alpha > 0$. Another is treating directly the long-run average per unit time. We shall show below two approaches. Let us consider the first approach. From (8.41) and (8.15), we have

$$(8.43) \qquad v(\alpha) = [I + m(\alpha)]\rho(\alpha).$$

We now assume that the semi-Markov process is ergodic. Then, using (8.32), the asymptotic behavior $m(\alpha)$ for a small α, and noting $\lim_{\alpha \to 0} \rho_i(\alpha) = r_i \eta_i$, we have

$$(8.44) \qquad v_i(\alpha) = \frac{g}{\alpha} + v_i + o(1), \qquad (i \in S)$$

where

$$(8.45) \qquad g=\sum_{j\in S}\frac{r_j\eta_j}{\mu_{jj}}=\sum_{j\in S}\pi_j r_j \eta_j \Big/ \sum_{j\in S}\pi_j \eta_j \,,$$

and

$$(8.46) \qquad v_i = r_i\eta_i + \sum_{j\in S} r_j\eta_j\left[\frac{\mu_{jj}^{(2)}}{2(\mu_{jj})^2} - \frac{\mu_{ij}}{\mu_{jj}} + \delta_{ij}\right].$$

Here we assume that $\mu_{jj}^{(2)}$ is finite. Note that g is the long-run average return per unit time and shows the result of the average of r_i ($i \in S$) with respect to the limiting probabilities in (8.27).

It is easy to extend the result of (8.44) to any ergodic state. For any ergodic state i belonging to an ergodic set $E_\nu \subset S$, we have the similar result (8.44) except that the summation is taken over E_ν instead of S. And that for (8.45) and (8.46) the summation is also taken over E_ν.

Applying the similar approach, we have for each state (see Fox [31])

$$(8.47) \qquad v(\alpha)=\frac{u}{\alpha} + v + o(1),$$

where u, v are $N \times 1$ column vectors. This corresponds to (7.62) in the discrete time case.

We shall show the second approach of treating directly the long-run average per unit time. Let

$\nu_i(t)$ be the total expected return up to time t starting in state i at time zero. Then $\nu_i(t)$ satisfies the following equation of renewal type in the same way as we have derived $P_{ij}(t)$. That is, we have

$$(8.48) \quad \nu_i(t) = [1 - H_i(t)]\, r_i t + \sum_{j \in S} \int_0^t [r_i \tau + \nu_j(t-\tau)]\, dQ_{ij}(\tau). \qquad (i \in S)$$

We shall first assume that a Markov renewal process is ergodic. In this case we shall consider only the average return per unit time in the steady state. Therefore, for a sufficiently large t, the first term of the right-hand side of (8.48) tends to zero because of the finiteness of the mean of $H_i(t)$ and of the fact that $[1 - H_i(t)]$ converges to zero more rapidly than $r_i t$ diverges. Consequently, for a sufficiently large t, we have

$$(8.49) \quad \nu_i(t) = \sum_{j \in S} \int_0^t [r_i \tau + \nu_j(t-\tau)]\, dQ_{ij}(\tau)$$
$$= \sum_{j \in S} \int_0^t r_i \tau\, dQ_{ij}(\tau) + \sum_{j \in S} \int_0^t \nu_j(t-\tau)\, dQ_{ij}(\tau). \qquad (i \in S)$$

The first term of the right-hand side of (8.49) tends to $r_i \eta_i$ as $t \to \infty$, then we have for a sufficiently large t

$$(8.50) \quad \nu_i(t) = r_i \eta_i + \sum_{j \in S} \int_0^t \nu_j(t-\tau)\, dQ_{ij}(\tau). \qquad (i \in S)$$

Let $v_i(s)$ be the Laplace transform of $v_i(t)$. We define two $N \times 1$ column vectors

$$(8.51) \qquad v(\alpha) = \begin{bmatrix} v_1(s) \\ \cdot \\ \cdot \\ \cdot \\ v_N(s) \end{bmatrix}, \qquad r = \begin{bmatrix} r_1\eta_1 \\ \cdot \\ \cdot \\ \cdot \\ r_N\eta_N \end{bmatrix}.$$

Applying the Laplace transforms to (8.50), we have

$$(8.52) \qquad v(s) = \frac{1}{s} r + \mathfrak{f}(s)\, v(s),$$

or

$$(8.53) \qquad v(s) = \frac{1}{s}[I - \mathfrak{f}(s)]^{-1} r.$$

Using (8.15), we have

$$(8.54) \qquad v(s) = \frac{1}{s}[I + m(s)]\, r.$$

The total expected return up to time t starting in an initial distribution a is $av(s)$, where $v(s)$ is the $N \times 1$ column vector with ith element $v_i(t)$. Since $\lim_{s\to 0} s^2 a[I + m(s)]r = \sum_{j\in S} (r_j\eta_j/\mu_{jj})$, from a Tauberian Theorem we have

$$(8.55) \qquad \lim_{t\to\infty} \frac{av(t)}{t} = \sum_{j\in S} \frac{r_j\eta_j}{\mu_{jj}} = \sum_{j\in S} \pi_j r_j \eta_j \Big/ \sum_{j\in S} \pi_j \eta_j,$$

which is independent of an initial distribution a. Thus g is the long-run average per unit time which has been obtained in (8.45).

We shall then consider the <u>terminating process</u> (see Section 7. 7). Since we consider only the total expected return before absorption, it is sufficient for this case that the semi-Markov process considered is absorbing. So, we assume that state 1 is absorbing and other states are transient. Let $v_i(t)$ be the total expected return up to time t. Then we have the following equation of renewal type in the same way as we have derived (8.48):

$$(8.56)\quad v_i(t)=[1-H_i(t)]r_i t+\int_0^t r_i\tau\, dQ_{i1}(t)$$
$$+\sum_{j=2}^{N}\int_0^t [r_i\tau+v_j(t-\tau)]\, dQ_{ij}(\tau). \qquad (i=2,\ldots,N)$$

While

$$(8.57)\qquad v_i=\lim_{t\to\infty} v_i(t) \qquad (i=2,\ldots,N)$$

which is valid since the system reaches the absorbing state with probability 1. Assuming $t\to\infty$ in (8.56), we have

$$(8.58)\quad v_i=\sum_{j\in S} r_i\int_0^\infty \tau\, dQ_{ij}(\tau)+\sum_{j=2}^{N} v_j\int_0^\infty dQ_{ij}(\tau)$$
$$=r_i\eta_i+\sum_{j=2}^{N} p_{ij}v_j \qquad (i=2,\ldots,N)$$

Introducing two $(N-1)\times 1$ column vectors

(8.59) $$\nu' = \begin{bmatrix} \nu_2 \\ \cdot \\ \cdot \\ \cdot \\ \nu_N \end{bmatrix}, \qquad r' = \begin{bmatrix} r_2\eta_2 \\ \cdot \\ \cdot \\ \cdot \\ r_N\eta_N \end{bmatrix},$$

we can rewrite (8.58) in matrix form as follows:

$$\nu' = r' + P_1 \nu',$$

where P_1 is the $(N - 1) \times (N - 1)$ submatrix of eliminating the first row and column. Thus we have

(8.60) $$\nu' = [I - P_1]^{-1} r',$$

where $[I - P_1]^{-1}$ is the fundamental matrix of the absorbing Markov chain in the discrete time case (see Kemeny and Snell [47, p.46]). Therefore, the total expected return before absorption starting in an initial distribution a is

(8.61) $$a'\nu' = a'[I - P_1]^{-1} r',$$

where $a' = (a_2, a_3, \ldots, a_N)$.

We have defined the return structure that we receive the return r_i when the system is in state i during a unit time. Now we shall consider the generalized returns (see Jewell [44]). When the system is in state i, and the next visiting state j, a return is accumulated according to the arbitrary function $R_{ij}(t \mid \tau)$, depending on i and j, the sojourn time $\tau(i, j)$, and on the clock time t since the beginning of the transition interval ($0 \leq t \leq \tau(i, j)$). We shall assume that $R_{ij}(0 \mid \tau) = 0$, and will denote the total return at the end of the interval by $R_{ij}(\tau \mid \tau) = R_{ij}(\tau)$. Returns from successive transitions are additive.

For the discounted model, the average, one-step-discounted return starting in state i is

$$(8.62) \qquad \rho_i(\alpha) = \sum_{j \in S} \int_0^\infty dQ_{ij}(\tau) \int_0^\tau e^{-\alpha x} d_x R_{ij}(x \mid \tau), \qquad (i \in S)$$

which corresponds to (8.38). Our return structure considered above is a special case $R_{ij}(x \mid \tau) = r_i x$. Thus the return structures are essentially the same ones for the discounted model.

For the undiscounted model we have in general $\rho_i(\alpha) \to \rho_i$ (constant) as $\alpha \to 0$, which corresponds to $\rho_i = r_i \eta_i$, but the bias term (8.46) will change slightly (see Jewell [45]).

Consequently, in both cases it suffices to consider only the simple return structures r_i ($i \in S$) instead of $R_{ij}(t \mid \tau)$.

8. 4. Markov Renewal Programs with Discounting

We shall consider Markov renewal programs with discounting using the results of the preceding section.

Markov renewal programs with which we are concerned may be stated as follows: When the system is in state i, we have K_i actions in each state $i \in S$. If we choose an action k in state $i \in S$, the system obeys the probability law

$Q_{ij}^{k}(t) = p_{ij}^{k} F_{ij}^{k}(t)$ $(j \in S)$, where p_{ij}^{k} is the transition probability from state i to state j and $F_{ij}^{k}(t)$ is the sojourn time in state i knowing that the next visiting state is j. And we get the return r_i^k when the system is in state i during a unit time. In other words, we have K_i selections in each state i, that is, we have $k = 1, \ldots, K_i$ actions

$$(8.63) \qquad Q_{ij}^{k}(t) \qquad (p_{ij}^{k},\ F_{ij}^{k}(t)) \qquad \text{and} \qquad r_i^k \quad (j \in S)$$

for each $i \in S$.

Then, what strategy, i.e., the sequence of actions in each transition epoch for each state over all $t \geq 0$, maximizes the total expected return (or the average return per unit time) starting in an initial distribution?

As we have seen in the preceding section, the total expected return is finite if the process is discounted or terminating, and is usually infinite if the process is nondiscounted. Thus we shall consider the total expected return if it is finite and the average return per unit time if it is infinite. In this section we shall first treat the discounted process.

Note that we can consider Markov renewal programs over a finite time horizon. But for the model the sojourn time in any state is a random variable and we cannot know the transition epoch. Thus we can expect no results so far as we know except the approximation approach.

For our model with an infinite time horizon we define a stationary strategy such that for each state we take the identical action independent of the history of previous states, transition times, and decisions. In general we may consider the nonstationary strategy depending on the history of states and transition times, and decisions.

Let $v(\alpha, \pi)$ be the discounted total expected return ($N \times 1$ column vector) starting in each state using any policy π, where a discount factor $\alpha > 0$.

Definition 8. 1. A strategy π^* is called α-optimal if $v(\alpha, \pi^*) \geq v(\alpha, \pi)$ for all π, where α ($\alpha > 0$) is fixed.

Then we have

Theorem 8. 2. There is an α-optimal strategy which is stationary.

The proof of this theorem has been established via contraction mapping (Denardo [18]).

From Theorem 8. 2 we may find α-optimal strategies within stationary strategies whose number is finite. Then we can use immediately the preceding result (8.41).

We now develop the linear programming formulation for Markov renewal program with discounting. From (8.41) we have for any stationary strategy

$$v(\alpha, \pi) = [I - q(\alpha)]^{-1} \rho(\alpha). \tag{8.64}$$

And the discounted total expected return starting in an initial distribution a is

$$a\, v(\alpha, \pi) = a\,[I - q(\alpha)]^{-1} \rho(\alpha) \tag{8.65}$$

$$= \sum_{i \in S} \sum_{j \in S} a_i \mu_{ij}(\alpha)\, \rho_j(\alpha),$$

where

$$[I - q(\alpha)]^{-1} = [\mu_{ij}(\alpha)], \tag{8.66}$$

and it holds that

$$[I - q(\alpha)]^{-1} [I - q(\alpha)] = I,$$

or

$$\sum_{j \in S} \mu_{ij}(\alpha)\, (\delta_{j\ell} - q_{j\ell}(\alpha)) = \delta_{i\ell}. \tag{8.67}$$

It is more convenient to extend of decisions to include randomized (or mixed) strategies. Let the probability that we make an action k when the system is in state i be d_i^k ($i \in S$, $k \in K_i$), which is independent of time t since we consider only the stationary strategies. It is clear that

$$d_i^k \geq 0, \quad \sum_{k \in K_i} d_i^k = 1. \qquad (i \in S, \ k \in K_i) \tag{8.68}$$

Using d_i^k, we have for any stationary strategy

(8.69) $$a\,v(\alpha,\pi)=\sum_{i\in S}\sum_{j\in S}\sum_{k\in K_j} a_i\,\mu_{ij}(\alpha)\,\rho_j^k(\alpha)\,d_j^k,$$

where

(8.70) $$\rho_j^k(\alpha)=\frac{r_i^k}{\alpha}\,[1-h_i^k(\alpha)],$$

(8.71) $$h_i^k(\alpha)=\sum_{j\in S}\int_0^\infty e^{-st}\,dQ_{ij}^k(t)=\sum_{j\in S} g_{ij}^k(\alpha).$$

Note that $\mu_{ij}(\alpha)$ depends on d_i^k since $[I - g(\alpha)]$ is determined by d_i^k. Now we define

(8.72) $$x_j^k=\sum_{i\in S} a_i\,\mu_{ij}(\alpha)\,d_j^k\geq 0.$$

Rewriting (8.69) by using x_j^k, we have

(8.73) $$a\,v(\alpha,\pi)=\sum_{j\in S}\sum_{k\in K_j}\rho_j^k(\alpha)\,x_j^k.$$

While we calculate

(8.74) $$\sum_{j\in S}\sum_{k\in K_j}(\delta_{j\ell}-g_{j\ell}^k(\alpha))x_j^k$$
$$=\sum_{i\in S}\sum_{j\in S}\sum_{k\in K_j}(\delta_{j\ell}-g_{j\ell}^k(\alpha))\,a_i\,\mu_{ij}(\alpha)\,d_j^k$$
$$=\sum_{i\in S}\sum_{j\in S} a_i\,\mu_{ij}(\alpha)\Big(\sum_{k\in K_j}(\delta_{j\ell}-g_{j\ell}^k(\alpha))\,d_j^k\Big)$$
$$=\sum_{i\in S} a_i\Big[\sum_{j\in S}\mu_{ij}(\alpha)\,(\delta_{j\ell}-g_{j\ell}(\alpha))\Big]$$

$$= \sum_{i \in S} a_i \delta_{i\ell} \qquad \text{(from (8.67))}$$

$$= a_\ell , \qquad (\ell \in S)$$

that is,

$$(8.75) \qquad \sum_{k \in K_j} x_j^k - \sum_{i \in S} \sum_{k \in K_i} \delta_{ij}^k(\alpha)\, x_i^k = a_j . \qquad (j \in S)$$

Thus we have the following linear programming problem:

$$(8.76) \qquad \text{Max} \quad \sum_{j \in S} \sum_{k \in K_j} \rho_j^k(\alpha)\, x_j^k$$

subject to

$$(8.77) \qquad \sum_{k \in K_j} x_j^k - \sum_{i \in S} \sum_{k \in K_i} \delta_{ij}^k(\alpha)\, x_i^k = a_j , \qquad (j \in S)$$

$$(8.78) \qquad x_j^k \geq 0. \qquad (j \in S, \quad k \in K_j)$$

This problem is similar to that of Section 6. 3. Here we apply the different derivation of Section 6. 3, but we can also use the similar approach of Section 6. 3. Thus we have the similar results of Section 6. 3 for this process. We don't state here repeatedly.

We show only the following theorem which is useful for the later discussion.

Theorem 8. 3. For all positive right-hand side $a_j > 0$ (say $a_j = 1/N$), there exists any basic feasible solution with property such that for each

$i \in S$ there is only one k such that $x_i^k > 0$ and $x_i^k = 0$ for k otherwise.

Since the proof is essentially the same one of Corollary 6. 11, we omit the proof.

This theorem asserts that any basic feasible solution forms a nonrandomized stationary strategy. This theorem plays an important role for the policy iteration algorithm.

We then shall consider the dual problem of the primal problem (8.76)-(8.78) (since the similar discussion of Section 6. 3 implies that the constraints (8.77) have rank N and the dual problem is meaning). The dual problem is:

(8.79) $$\text{Min} \sum_{i \in S} a_i v_i$$

subject to

(8.80) $$v_i \geq r_i^k(\alpha) + \sum_{j \in S} q_{ij}^k(\alpha) v_j, \qquad (i \in S, \; k \in K_i)$$

(8.81) v_i; unconstrained in sign for $i \in S$,

where the dual variable v_i ($i \in S$) is the ith element of $v(\alpha, \pi^*)$, where π^* is an α-optimal strategy.

Fig. 3.1 shows the Tucker diagram of the primal and dual problems. The relation between policy iteration and linear programming algorithms has been stated in Section 6. 4. For this problem we can apply the similar approach. Using the results of Section 6. 4 for this problem and Theorem 8. 3 of the

Dual	Primal									
	$x_1^1 \geq 0$	$x_1^2 \geq 0$	$\cdots$	$x_2^1 \geq 0$	$x_2^2 \geq 0$	$\cdots$	$x_N^1 \geq 0$	$x_N^2 \geq 0$ $\cdots$	Relations	Variables
v_1	$1-\gamma_{11}^1(\alpha)$	$1-\gamma_{11}^2(\alpha)$	$\cdots$	$-\gamma_{21}^1(\alpha)$	$-\gamma_{21}^2(\alpha)$	$\cdots$	$-\gamma_{N1}^1(\alpha)$	$-\gamma_{N1}^2(\alpha)$ $\cdots$	$=$	a_1
v_2	$-\gamma_{12}^1(\alpha)$	$-\gamma_{12}^2(\alpha)$	$\cdots$	$1-\gamma_{22}^1(\alpha)$	$1-\gamma_{22}^2(\alpha)$	$\cdots$	$-\gamma_{N2}^1(\alpha)$	$-\gamma_{N2}^2(\alpha)$ $\cdots$	$=$	a_2
$\cdot$	$\cdot$	$\cdot$		$\cdot$	$\cdot$		$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$		$\cdot$	$\cdot$		$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$		$\cdot$	$\cdot$		$\cdot$	$\cdot$	$\cdot$	$\cdot$
v_{N-1}	$-\gamma_{1,N-1}^1(\alpha)$	$-\gamma_{1,N-1}^2(\alpha)$	$\cdots$	$-\gamma_{2,N-1}^1(\alpha)$	$-\gamma_{2,N-1}^2(\alpha)$	$\cdots$	$-\gamma_{N,N-1}^1(\alpha)$	$-\gamma_{N,N-1}^2(\alpha)$ $\cdot$	$=$	a_{N-1}
v_N	$-\gamma_{1N}^1(\alpha)$	$-\gamma_{1N}^2(\alpha)$	$\cdots$	$-\gamma_{2N}^1(\alpha)$	$-\gamma_{2N}^2(\alpha)$	$\cdots$	$1-\gamma_{NN}^1(\alpha)$	$1-\gamma_{NN}^2(\alpha)$ $\cdot$	$=$	a_N
Relations	$\geq$	$\geq$	$\cdots$	$\geq$	$\geq$	$\cdots$	$\geq$	$\geq$ $\cdots$		
Constants	$\rho_1^1(\alpha)$	$\rho_1^2(\alpha)$	$\cdots$	$\rho_2^1(\alpha)$	$\rho_2^2(\alpha)$	$\cdots$	$\rho_N^1(\alpha)$	$\rho_N^2(\alpha)$ $\cdots$		

Fig. 8. 1. The Tucker diagram for the semi-Markovian decision process with discounting.

primal problem, we have immediately the policy iteration algorithm for Markov renewal program with discounting. We first note that the dual variables in (8.81) are also the simplex multipliers for the primal problem.

Policy iteration algorithm is the following two parts.

Value Determination Operation

Take any stationary strategy f^{∞}. Solve

$$v_i = \rho_i^k(\alpha) + \sum_{j \in S} q_{ij}^k(\alpha) v_j$$

for v_i ($i \in S$), where the superscript k corresponds to the chosen strategy f^{∞}.

Policy Improvement Routine

Using the values v_i($i \in S$), find the element of $G(i, f)$ for each $i \in S$ such that

$$\rho_i^k(\alpha) + \sum_{j \in S} q_{ij}^k(\alpha) v_j > v_i$$

for all $k \in K_i$. If $G(i, f)$ is empty for all $i \in S$; f^{∞} is α-optimal and $v(\alpha , f^{\infty}) = [v_i]$ is the discounted total expected return. If at least $g(i) \in G(i , f)$ for some i, make an improved strategy g^{∞} such that $g(i) \in G(i , f)$ for some i and $g(i) = f(i)$ for $G(i , f)$ empty, and return to Value Determination Operation.

Here we use the same notation f^∞, which has denoted a stationary strategy in the discrete time case.

As an initial strategy f^∞, we may take, for example, $\max_{k \in K_i} \rho_i^k(\alpha)$ for each $i \in S$, the reason of this selection can be seen from the linear programming viewpoint.

From the linear programming viewpoint, Policy Improvement Routine is an extension of the simplex criterion such that the simplex criteria for each state are applied simultaneously. If $g(i) \in G(i, f)$ for some one i and $G(i, f)$ is empty for other i, Policy Improvement Routine is equivalent to the simplex criterion of the primal problem.

We shall show that an improved strategy g^∞ from Policy Improvement Routine satisfies $v(\alpha, g^\infty) > v(\alpha, f^\infty)$.

Theorem 8. 4. Take any f^∞. If $g(i) \in G(i, f)$ for some i and otherwise $g(i) = f(i)$ for $G(i, f)$ empty, then $v(\alpha, g^\infty) > v(\alpha, f^\infty)$.

Proof. For two strategies f^∞ and g^∞, we have

(8.82) $$v(\alpha, f^\infty) = \rho^f(\alpha) + q^f(\alpha)\, v(\alpha, f^\infty),$$

(8.83) $$v(\alpha, g^\infty) = \rho^g(\alpha) + q^g(\alpha)\, v(\alpha, f^\infty).$$

Substracting (8.82) from (8.83), we have

(8.84) $$v(\alpha, g^\infty) - v(\alpha, f^\infty) = \rho^g(\alpha) - \rho^f(\alpha) + q^g(\alpha) v(\alpha, g^\infty) - q^f(\alpha) v(\alpha, f^\infty).$$

We define the $N \times 1$ column vector

$$(8.85)\quad \gamma = [\gamma_i] = \rho^g(\alpha) - \rho^f(\alpha) + \mathfrak{g}^g(\alpha)\, v(\alpha, g^\infty) - \mathfrak{g}^f(\alpha)\, v(\alpha, f^\infty).$$

where $\gamma_i > 0$ if $g(i) \in G(i, f)$ for some i and $\gamma_i = 0$ if $G(i, f)$ is empty from the hypothesis of the theorem. Combining (8.84) and (8.85), and defining $\Delta v = v(\alpha, g^\infty) - v(\alpha, f^\infty)$, we have

$$(8.86)\qquad \Delta v = \gamma + \mathfrak{g}^g(\alpha)\,\Delta v.$$

Solving for Δv, we obtain

$$(8.87)\qquad \Delta v = [I - \mathfrak{g}^g(\alpha)]^{-1}\gamma.$$

Since (8.85) and $[I - \mathfrak{g}^g(\alpha)]^{-1}$ is nonnegative and these are never zero simultaneously, we have that at least one element of Δv is positive, i.e., $\Delta v > 0$, which completes the proof.

We have seen from Theorem 8. 4 that the policy iteration algorithm terminates an α-optimal stationary strategy with finite iterations, since the number of stationary strategies is finite. Note that if there are two or more strategies which satisfy the optimality equation $v(\alpha, f^\infty) = \max_f[\rho^f(\alpha) + \mathfrak{g}^f(\alpha)\, v(\alpha, f^\infty)]$, then these strategies are all α-optimal.

8. 5. Markov Renewal Programs with No Discounting

Now we shall develop the semi-Markovian decision processes wiht no discounting. For the discounted process we have seen that there is an α-optimal stationary strategy. For the nondiscounted process we shall first prove the same result.

As a criterion for the nondiscounted model, we define for any strategy

$$\mathcal{U}(\pi) = \liminf_{t\to\infty} \frac{v(t;\pi)}{t}, \tag{8.88}$$

where $v(t;\pi)$ is the total expected return ($N \times 1$ column vector) up to time t starting in each state using a strategy π. Then $u(\pi)$ is the long-run average return per unit time, which corresponds to (6.12) in the discrete time case. If we take a stationary strategy, we have $v(t;\pi)$ by solving (8.48) for $v_i(t)$. But for any nonstationary strategy, we cannot write explicitly.

Our problem is then to find a strategy which maximizes (8.88) under all strategies π, i.e., an optimal strategy π^* such that $\mathcal{U}(\pi^*) \geq \mathcal{U}(\pi)$ for all π. The following theorem has been given by Fox [31].

Theorem 8. 5. There is an optimal strategy which is stationary.

Proof. We have seen from Theorem 8. 2 that there is an α-optimal stationafy strategy. Since the number of stationary strategies is finite, then it is possible to choose a sequence $\{\alpha_\nu\}$, $\lim_{\nu\to\infty} \alpha_\nu = 0$, such that

$\pi^* = \pi_{\alpha_\nu}$, $\nu = 1, 2, \ldots$, where π_{α_ν} is a stationary strategy. While, we have from (8.47) that

$$(8.89) \qquad v(\alpha, f) = \frac{u(f)}{\alpha} + v(f) + o(1)$$

for any stationary strategy f^∞. Thus using Abelian and Tauberian Theorems, we have

$$(8.90) \; u(\pi) \leq \liminf_{\nu \to \infty} \alpha_\nu v(\alpha_\nu, \pi) \leq \liminf_{\nu \to \infty} \alpha_\nu v(\alpha_\nu, \pi_{\alpha_\nu}) = u(\pi^*)$$

for any strategy π, which shows that π^* is an optimal stationary strategy. Thus we complete the proof.

Applying Theorem 8. 5, we may find optimal strategies within the stationary ones under the average criterion. Thus we restrict our attention to stationary strategies and then we can immediately apply the results of Section 8. 3.

In the remaining part of this section we shall discuss the linear programming formulation for the completely ergodic process and the terminating process. Further we can obtain similarly the corresponding policy iteration algorithms from the straightforward results of Chapter VII.

First we shall consider the completely ergodic case. For this problem the long-run average return per unit time is

$$(8.91) \qquad g = \sum_{j \in S} \pi_j \eta_j r_j \Big/ \sum_{j \in S} \pi_j \eta_j ,$$

where π_j is the limiting probability in state j of the imbedded Markov chain. Now we shall consider the decision process. Then the objective function is

$$(8.92) \qquad \sum_{j\in S} \pi_j(f)\eta_j^k r_j^k \Big/ \sum_{j\in S} \pi_j(f)\eta_j^k ,$$

where $\pi_j(f)$ is the corresponding limiting probability of a stationary strategy f^∞, and η_i^k, r_i^k are the corresponding unconditional mean, the return of the strategy f^∞, respectively. The limiting probabilities satisfy

$$(8.93) \qquad \pi_j(f) = \sum_{i\in S} \pi_i(f)\, p_{ij}(f), \qquad (j\in S)$$

$$(8.94) \qquad \sum_{j\in S} \pi_j(f) = 1,$$

$$(8.95) \qquad \pi_j(f) > 0. \qquad (j \in S)$$

Then, let d_j^k ($j\in S$, $k\in K_j$) be the joint probability that the system is in state i and the decision k is made. It is evident that

$$(8.96) \qquad \sum_{k\in K_j} d_j^k = 1, \qquad 0 < d_j^k (\leq 1). \qquad (j\in S,\ k\in K_j)$$

The objective function of our problem is

$$(8.97) \qquad \sum_{j\in S}\sum_{k\in K_j} \pi_j(f)\eta_j^k r_j^k d_j^k \Big/ \sum_{j\in S}\sum_{k\in K_j} \pi_j(f)\eta_j^k d_j^k$$

by using d_j^k. While (8.93), (8.94) and (8.95) become

(8.98) $\pi_j(f) - \sum_{i\in S}\sum_{k\in K_i} \pi_i(f)\, p_{ij}^k d_i^k = 0,$ $\qquad (j \in S)$

(8.99) $\sum_{j\in S} \pi_j(f) = 1,$

(8.100) $\pi_j(f) > 0.$ $\qquad (j \in S)$

Setting

(8.101) $x_j^k = \pi_j(f)\, d_j^k \geq 0,$ $\qquad (j \in S,\ k \in K_j)$

and using the fact that $\pi_j(f) = \sum_{k\in K_j} x_j^k$ for $j \in S$, we have from (8.97)-(8.101) the following programming problem:

(8.102) Max $\sum_{j\in S}\sum_{k\in K_j} \eta_j^k r_j^k x_j^k \Big/ \sum_{j\in S}\sum_{k\in K_j} \eta_j^k x_j^k$

subject to

(8.103) $\sum_{k\in K_j} x_j^k - \sum_{i\in S}\sum_{k\in K_i} p_{ij}^k x_i^k = 0,$ $\qquad (j \in S)$

(8.104) $\sum_{j\in S}\sum_{k\in K_j} x_j^k = 1,$

(8.105) $x_j^k \geq 0.$ $\qquad (j \in S,\ k \in K_j)$

This problem is the so-called <u>linear fractional programming</u> problem (see, e.g., Charnes and Cooper [12]) since the objective function (8.102) is linear fractional and the constraints are linear. So we can reduce to a linear programming problem by using variables transformation. Noting that $\sum_{j\in S}\sum_{k\in K_j} \eta_i^k x_i^k > 0,$

we transform the variables x_j^k to y_j^k and y such that

(8.106) $$y_j^k = x_j^k \Big/ \sum_{j \in S} \sum_{k \in K_j} \eta_j^k x_j^k ,$$

and

(8.107) $$y = 1 \Big/ \sum_{j \in S} \sum_{k \in K_j} \eta_j^k x_j^k .$$

Then we have the following linear programming problem:

(8.108) Max $$\sum_{j \in S} \sum_{k \in K_j} \eta_j^k r_j^k y_j^k$$

subject to

(8.109) $$\sum_{k \in K_j} y_j^k - \sum_{i \in S} \sum_{k \in K_i} p_{ij}^k y_i^k = 0, \qquad (j \in S)$$

(8.110) $$\sum_{j \in S} \sum_{k \in K_j} y_j^k = y ,$$

(8.111) $$\sum_{j \in S} \sum_{k \in K_j} \eta_j^k y_j^k = 1,$$

(8.112) $$y_j^k \geq 0, \qquad (j \in S , \quad k \in K_j)$$

where the additional constraint (8.111) is derived from the transforms of (8.106) and (8.107).

Now we shall consider some properties of the above linear programming problem.

Theorem 8. 6. For the linear programming problem (8.108)-(8.112), there is any basic feasible solution with property that, for each $i \in S$, there is only one k such that $y_i^k \geq 0$ and $y_i^k = 0$ for k otherwise.

The proof of this theorem is similar to that of Theorem 7. 8. So we omit the proof here.

Theorem 8. 6 implies that any basic feasible solution has always $y > 0$.

Theorem 8. 7. Semi-Markov decision process under consideration is equivalent to the linear programming problem:

(8.113) $\quad$ Max $\displaystyle\sum_{j \in S} \sum_{k \in K_j} \eta_j^k r_j^k y_j^k$

subject to

(8.114) $\quad \displaystyle\sum_{k \in K_j} y_j^k - \sum_{i \in S} \sum_{k \in K_i} p_{ij}^k y_i^k = 0, \qquad (j = 1, \ldots, N-1)$

(8.115) $\quad \displaystyle\sum_{j \in S} \sum_{k \in K_j} \eta_j^k y_j^k = 1,$

(8.116) $\quad y_j^k \geq 0. \qquad (j \in S, \; k \in K_j)$

That is, we need not to take account of the constraint (8.110) and of the variable y. Here we omit the redundant constraint for $j = N$ in (8.114).

Proof. As we have noted above, any basic feasible solution has always $y > 0$. But we should require the optimal stationary and its associated maximum average return per unit time g. It is clear that the optimal value of (8.113) has no relation to y. From the transforms of variables in (8.106) and (8.107), we have $x_j^k = y y_j^k$ for $j \in S$, $k \in K_j$. Thus the optimal stationary strategy is determined by

$$(8.117) \qquad d_j^k = \frac{x_j^k}{\sum_{k \in K_j} x_j^k} = \frac{y y_j^k}{\sum_{k \in K_j} y y_j^k} = \frac{y_j^k}{\sum_{k \in K_j} y_j^k},$$

which is independent of y. These completes the proof.

Theorem 8. 7 gives the linear programming problem (8.113)-(8.116) which has $\sum_{i \in S} K_i$ variables and N constraints, where the redundant sonstraint for $j = N$ in (8.114) is omitted.

Now we shall consider the dual problem (8.113)-(8.116) as a primal problem. Defining dual variables ($v_1, v_2, \ldots, v_{N-1}, g$), we have the dual problem:

$$(8.118) \qquad \text{Min} \quad g$$

subject to

$$(8.119) \qquad \eta_i^k g + v_i \geq \eta_i^k r_i^k + \sum_{j=1}^{N-1} p_{ij}^k v_j , \qquad (j \in S,\ k \in K_j)$$

(8.120) $\quad g$, v_i; unconstrained in sign for $i = 1, 2, \ldots, N-1$.

where we suppose $\nu_N = 0$ in (8.119). Here we have used the known fact that the constraints (8.114) and (8.115) have rank N and the dual problem is meaning. Further we have also used the fact that the objective function of the dual problem is g, the average return per unit time, from the duality theorem (see, e.g., Dantzig [15]). Fig. 8. 2 shows the Tucker diagram of the primal and dual problems for the completely ergodic semi-Markovian decision process.

We see the similar discussion of Section 7. 4 that the dual variables ν_i ($i = 1, 2, \ldots, N-1$) are relative bias terms, which correponds to (8.46). For this problem we can obtain the policy iteration algorithm from the similar discussion of Section 7. 4. We note that the dual variables ($\nu_1, \nu_2, \ldots, \nu_{N-1}, g$) are also the simplex multipliers. From the linear programming viewpoint we have for any basic variables

$$(8.121) \quad \Delta_i^* = -\eta_i^* r_i^* - \sum_{j=1}^{N-1} p_{ij}^* \nu_j - \eta_i^* g - \nu_i = 0, \qquad (i \in S)$$

where the asterisk $*$ denotes the data of any basic feasible solution (and also any stationary strategy). While, if

$$(8.122) \quad \Delta_i^k = -\eta_i^k r_i^k - \sum_{j=1}^{N-1} p_{ij}^k \nu_j + \eta_i g + \nu_i < 0$$

for some $i \in S$ and $k \in K_i$, we can obtain an improved strategy. Further if $\Delta_i^k \geq 0$ for all $i \in S$ and $k \in K_i$, we have an optimal strategy. Solving for g in (8.122),

Dual \ Primal	$y_1^1 \geq 0$	$y_1^2 \geq 0$	$\cdots$	$y_2^1 \geq 0$	$y_2^2 \geq 0$	$\cdots$	$y_N^1 \geq 0$	$y_N^2 \geq 0$	$\cdots$	Relations	Variables
v_1	$1-p_{11}^1$	$1-p_{11}^2$	$\cdots$	$-p_{21}^1$	$-p_{21}^2$	$\cdots$	$-p_{N1}^1$	$-p_{N1}^2$	$\cdots$	$=$	0
v_2	$-p_{12}^1$	$-p_{12}^2$	$\cdots$	$1-p_{22}^1$	$1-p_{22}^1$	$\cdots$	$-p_{N2}^1$	$-p_{N2}^2$	$\cdots$	$=$	0
$\cdot$	$\cdot$	$\cdot$		$\cdot$	$\cdot$		$\cdot$	$\cdot$		$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$		$\cdot$	$\cdot$		$\cdot$	$\cdot$		$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$		$\cdot$	$\cdot$		$\cdot$	$\cdot$		$\cdot$	$\cdot$
v_{N-1}	$-p_{1,N-1}^1$	$-p_{1,N-1}^2$	$\cdots$	$-p_{2,N-1}^1$	$-p_{2,N-1}^2$	$\cdots$	$-p_{N,N-1}^1$	$-p_{N,N-1}^2$	$\cdots$	$=$	0
g	η_1^1	η_1^2	$\cdots$	η_2^1	η_2^2	$\cdots$	η_N^1	η_N^2	$\cdots$	$=$	1
Relations	$\geq$	$\geq$	$\cdots$	$\geq$	$\geq$	$\cdots$	$\geq$	$\geq$	$\cdots$		
Variables	$\eta_1^1 r_1^1$	$\eta_1^2 r_1^2$	$\cdots$	$\eta_2^1 r_2^1$	$\eta_2^2 r_2^2$	$\cdots$	$\eta_N^1 r_N^1$	$\eta_N^2 r_N^2$	$\cdots$		

Fig. 8. 2. The Tucker disgram for the completely ergodic semi-Markovian decision process.

we have

(8.123) $$g < r_i^k + \frac{1}{\eta_i^k}\left[\sum_{j=1}^{N-1} p_{ij}^k v_j - v_i\right].$$

which implies Policy Improvement Routine in the policy iteration algorithm. Thus in the above discussion we have the following policy iteration algorithm.

Value Determination Operation

Take any stationary f^∞. Solve

$$\eta_i^k g + v_i = \eta_i^k r_i^k + \sum_{j=1}^{N-1} p_{ij}^k v_j$$

for $g, v_1, v_2, \ldots, v_{N-1}$ (setting $v_N = 0$), where the superscript k is determined by the chosen strategy f^∞.

Policy Improvement Routine

Using the values v_i, find the element of $G(i, f)$ for each $i \in S$ such that

$$r_i^k + \frac{1}{\eta_i^k}\left[\sum_{j=1}^{N-1} p_{ij}^k v_j - v_i\right] > g$$

for all $k \in K_i$. If $G(i, f)$ is empty for all $i \in S$, f^∞ is optimal and g is the average return per unit time, $v_1, \ldots, v_{N-1}$ are the relative bias terms. If at least $g(i) \in G(i, f)$ for some i, make an improved strategy such that $g(i) \in G(i, f)$ for some i and $g(i) = f(i)$ for $G(i, f)$ empty, and return to Value Determination Operation.

We shall only show that the Policy Improvement Routine yields the improved strategy whose average return is greater than the previous one.

<u>Theorem 8. 8.</u> Take any f^{∞}. If $g(i) \in G(i, f)$ for some i and otherwise $g(i) = f(i)$ for $G(i, f)$ empty, then $g(g) > g(f)$, where $g(f)$ denotes the average return per unit time using a strategy f^{∞}.

<u>Proof.</u> For any two strategies f^{∞} and g^{∞}, we have

$$(8.124) \qquad \eta_i^f g(f) + v_i(f) = \eta_i^f r_i^f + \sum_{j=1}^{N-1} p_{ij}^f v_j(f),$$

$$(8.125) \qquad \eta_i^g g(g) + v_i(g) = \eta_i^g r_i^g + \sum_{j=1}^{N-1} p_{ij}^g v_j(g),$$

where η_i^f, p_{ij}^f denote η_i^k, p_{ij}^k, respectively, where $k = f(i)$. Dividing (8.124), (8.125) by η_i^f, η_i^g, respectively, and subtracting the former from the latter, we have

$$(8.126) \qquad \frac{v_i(g)}{\eta_i^g} - \frac{v_i(f)}{\eta_i^f} + g(g) - g(f) = r_i^g - r_i^f + \frac{1}{\eta_i^g}\sum_{j=1}^{N-1} p_{ij}^g v_j(g) - \frac{1}{\eta_i^f}\sum_{j=1}^{N-1} p_{ij}^f v_j(f).$$

We define the $N \times 1$ column vector

$$(8.\ 127) \qquad \gamma = [\gamma_i] = \left[r_i^g - r_i^f - \frac{1}{\eta_i^g}\left\{\sum_{j=1}^{N-1} p_{ij}^g v_j(g) - v_i(g)\right\} - \frac{1}{\eta_i^f}\left\{\sum_{j=1}^{N-1} p_{ij}^f v_j(f) - v_i(f)\right\}\right],$$

where $\gamma_i > 0$ if $g(i) \in G(i, f)$ for some i and $\gamma_i = 0$ if $G(i, f)$ is empty from the hypothesis of the theorem. Note that at least one $\gamma_i > 0$ for some i. Combining (8.126) and (8.127) and defining $\Delta v_i = v_i(g) - v_i(f)$ ($i = 1, \ldots, N-1$), $\Delta g = g(g) - g(f)$, we have

$$\Delta v_i + \eta_i^g \Delta g = \eta_i^g \gamma_i + \sum_{j=1}^{N-1} p_{ij}^g \Delta v_j. \tag{8.128}$$

Let $\pi(g) = (\pi_1(g), \ldots, \pi_N(g))$ be the limiting probability of the imbedded Markov chain $P(g) = [p_{ij}^g]$. It is evident that $\pi(g) = \pi(g) P(g)$. Premultiplying (8.128) by $\pi_i(g)$ and summing on all $i \in S$, we have

$$\Delta g = \sum_{i \in S} \pi_i(g) \eta_i^g \gamma_i \Big/ \sum_{i \in S} \pi_i(g) \eta_i^g, \tag{8.129}$$

which is positive since the limiting probability in state i is $\pi_i(g) \eta_i^g / \sum_{i \in S} \pi_i(g) \eta_i^g > 0$ for all $i \in S$, and at least one $\gamma_i > 0$. Thus we have

$$\Delta g = g(g) - g(f) > 0, \tag{8.130}$$

which completes the proof.

Theorem 8. 8 implies that the policy iteration algorithm terminates an optimal stationary strategy with finite iterations. Note that if there are two or more strategies which satisfy the <u>optimality equation</u> $\eta_i^k g + v_i = \max [\eta_i^k r_i^k + \sum_{j=1}^{N-1} p_{ij}^k v_j]$ ($i \in S$), then these strategies are all optimal and have the same average return g.

In this case we encounter the problem for finding a strategy having the maximal bias term among the optimal strategies. That is, we should require a 0-optimal strategy which refers to a 1-optimal strategy in the discrete time model. We omit the problem for finding 0-optimal strategies here.

Finally we shall consider the terminating process. For the terminating process the total expected return before absorption starting in an initial distribution is

$$\alpha' v' = \alpha' [I - P_1]^{-1} r' \tag{8.131}$$

where we define that state 1 is absorbing. For the semi-Markovian decision process with which we are concerned we shall first provide the following assumption (confer Section 7. 6).

Terminating Assumption. The common absorbing state is reachable with probability 1 from any transient state in a finite time whatever decisions we make.

Under the above assumption our problem is to find a strategy having the maximal total expected return before absorption among all strategies, thus we need not to consider the decision of state 1 since our concern is the behavior before absorption. This problem is very similar to that of the discounted model. So we shall only give the following linear programming problem:

(8.132) $$\text{Max} \quad \sum\sum r_j^k \eta_i^k x_j^k$$

subject to

(8.133) $$\sum_{k\in K_j} x_j^k - \sum_{i\in S}\sum_{k\in K_i} p_{ij}^k x_i^k = a_j , \qquad (j = 2, \ldots, N)$$

(8.134) $$x_j^k \geq 0. \qquad (j = 2,\ldots, N;\ k\in K_j)$$

The next theorem is obvious.

Theorem 8. 9. For all positive right-hand side $a_j > 0$ (say $a_j = 1/(N-1)$), there exists any basic feasible solution with property such that for each $i = 2,\ldots, N$, there is only k such that $x_i^k > 0$ and $x_i^k = 0$ for k otherwise.

The above linear programming problem and Theorem 8. 9 imply immediately the corresponding policy iteration algorithm. Here we don't state the algorithm.

CONCLUSION

9. 1. Summary of the Results

We summarize the results of the thesis in this section. As titled in the thesis, Markov renewal processes and their associated processes such as renewal processes and discrete time Markov chains are treated throughout this thesis. A Markov renewal process is one of the most powerful mathematical tools for analyzing and synthesizing systems.

In the first part of this thesis (Chapters I-V), we discussed some of redundant repairable systems. Some redundant systems treated in this thesis are of importance in the practical fields.

Mathematical techniques throughout the first part of this thesis are applications of Markov renewal processes (which include renewal processes). Taking account of the regeneration points of the failure (or the inspection) time distributions of units, we obtained the required Laplace-Stieltjes transforms. Systems considered in this thesis were a two-unit standby redundant system, a two-unit paralleled redundant system, an m-out-of-n system, and the modified systems just mentioned above.

Signal flow graph representation of systems is of great interest for engineers. We cannot understand at a glance the behavior of a large-scale and and complicated system. The signal flow graph method makes us suggestive and helpful, and the required quantities can be obtained by using Mason's gain formula which is a mechanical procedure. Examples of system analysis in Chapters IV and V are applications of the signal flow graph method.

In the second part of this thesis (Chapters VI-VIII), we discussed Markovian decision processes and Markov renewal programs as system synthesis. Markovian decision processes are one of the recent topics in Operations Research and Mathematical Statistics.

Policy iteration and linear programming algorithms were described and the relationship between the two algorithms was discussed in the viewpoint of Mathematical Programming. The discount factor in Markovian decision processes (and also in Markov renewal programs) plays an important role in analysis. For the discounted model we can analyse the processes elegantly since the processes have simple structures.

For the nondiscounted model we must consider the average return criterion if the total expected return is divergent. For the model a Markov chain considered changes its state classification from strategy to strategy. The policy iteration algorithm for the nondiscounted model was discussed. Linear programming considerations for the nondiscounted model were studied in detail and the relationship between the two algorithms made clear.

Markov renewal processes were reviewed in Chapter VIII. Markov renewal processes with returns were discussed for the discounted and the nondiscounted models. Markov renewal programs with discounting and with no discounting were discussed, and the policy iteration and linear programming algorithms were studied.

9. 2. Further Problems of System Analysis and Synthesis

We shall describe the further problems of system analysis and synthesis. The problems of system analysis and synthesis discuused in this thesis have the possibilities of extensions and modifications of models. We shall simply discuss the possibilities and suggestive comments.

Reliability analysis of redundant repairable systems has many fruitful studies. This thesis discussed only simple models and we restricted out attetion to the first passage times. In the actual situations we should consider more complicated models and futher discuss the mixed configurations of the models. Our concerns are also extended to not only the first passage times but also the transition probabilities, the limiting probabilities, etc.

We did not consider the factors of costs, weights, capacites, etc., which were associated with the models, for the analsysis of redundant systems. In the actual situations such factors will be imposed. We should consider the optimization problems of attaining the maximal reliability subject to the suitable constraints on such factors.

Markovian decision processes have also many problems and applications. Though we omitted in this thesis, 1-optimal strategies are of interest and many contributions to 1-optimal strategies have been made and will be made in future. Markov renewal programs have also the similar problems concerning 0-optimal strategies which correspond to 1-optimal ones in discrete time case.

Applications of Markovian decision processes and Markov renewal programs are of great interest. Markovian replacement problems have been disccused by many authors and will be discussed in futute. Applications to sampling inspection plan, quality control, reject allowance problem are also of interest.

9. 3. Publications List of the Author

[1] Mine, H. and Osaki, S., and Asakura, T., "Reliability considerations on redundant systems with repair," Memoirs of the Faculty of Engineering, Kyoto University, Kyoto, vol. 29 (1967), pp. 509-529.

[2] Mine, H. and Osaki, S., "On the reliability of parallel redundant systems with repair," (in Japanese) J. of Japan Assoc. of Automatic Control Engineers, vol. 12 (1968), pp. 265-271.

[3] Osaki, S. and Mine, H., "Linear programming algorithms for semi-Markovian decision processes," J. Math. Anal. Appl., vol. 22 (1968), pp. 356-381.

[4] Osaki, S. and Mine, H., "Some remarks on a Markovian decision problem with an absorbing state," Ibid., vol. 23 (1968), pp. 327-333.

[5] Mine, H., Osaki, S., and Asakura, T., "Some considerations for multiple-unit redundant systems with generalized repair time distributions," IEEE Trans. on Reliability, vol. R-17 (1968), pp. 171-174.

[6] Osaki, S. and Mine, H., "Linear programming considerations on Markovian decision processes with no discounting," J. Math. Anal. Appl., vol. 26 (1969), pp. 221-232.

[7] Mine, H. and Osaki, S., "On the reliability of multiple unit systems," (in Japanese) J. of Japan Assoc. of Automatic Control Engineers, vol. 13 (1969), pp. 271-277.

[8] Mine, H., Osaki, S., and Asakura, T., "On the reliability of multiple unit systems," (in Japanese) Trans. of Inst. of Elec. and Commun. Eng. of Japan, vol. 52-C (1969), pp. 241-242.

[9] Mine, H. and Osaki, S., "A note on a standby redundant system with noninstantaneous switchover," (in Japanese) Ibid., vol. 52-C (1969), pp. 437-438.

[10] Mine, H. and Osaki, S., "Some reliability aspects of complex systems," Proc. International Conference on Quality Control, 1969-Tokyo, pp. 197-200, Oct. 21-23, Tokyo, 1969.

[11] Mine, H. and Osaki, S., "On failure time distributions for systems of dissimilar units," IEEE Trans. on Reliability, vol. R-18 (1969), to appear.

[12] Osaki, S., "A standby redundant system with standby failure,"(in Japanese) Trans. of Inst. of Elec. and Commun. Eng. of Japan, vol. 52-C (1969), to appear.

[13] Mine, H., Osaki, S., and Asakura, T., "On a two unit standby redundant system and its maintenance," (in Japanese) Keiei-Kagaku (Official J. of Operations Research Soc. of Japan), to appear.

[14] Osaki, S., "Reliability analysis of a two-unit standby redundant system with standby failure," Opsearch (Official J. of Operational Research Soc. of India), vol.7 (1970), to appear.

[15] Mine, H., Yamada, K., and Osaki, S., "On terminating stochastic games," Management Sci., to appear.

[16] Osaki, S., "Reliability analysis of two-unit redundant systems," (in Japanese) Keiei-Kagaku, to appear.

[17] Osaki, S., "Reliability analysis of a two-unit standby redundant system with priority," J. of Canadian Operational Research Soc., vol. 8 (1970), to appear.

[18] Osaki, S., "A note on a two-unit standby redundant system," J. of Operations Research Soc. of Japan, to appear.

[19] Mine, H. and Osaki, S., Markovian Decision Processes, American Elsevier Publishing, Inc., Modern Analytic and Computational Methods in Sciences and Mathematics, No. 25, New York, 1970 (in press).

REFERENCES

[1] Barlow, R. E., "Repairman problems," in Studies in Applied Probability and Management Science, edited by Arrow, Karlin, and Scarf, Chap. 2, Stanford University Press, Stanford, 1962.

[2] ________, "Applications of semi-Markov processes to counter problems," Ibid., Chap. 3, 1962.

[3] ________ and Hunter, L. C., "Optimum preventive maintenance policies," Operations Research, vol. 8 (1960), pp. 90-100.

[4] ________ and Proschan, F., "Planned replacement," in Studies in Applied Probability and Management Science, edited by Arrow, Karlin, and Scarf, Chap. 4, Stanford University Press, Stanford, 1962.

[5] ________ and ________, Mathematical Theory of Reliability, Wiley, New York, 1965.

[6] Bellman, R., "A Markovian decision process," J. Math. Mech., vol. 6 (1957), pp. 679-684.

[7] ________, Dynamic Programming, Princeton University Press, Princeton, 1957.

[8] ________ and Dreyfus, S., Applied Dynamic Programming, Princeton University Press, Princeton, 1962.

[9] Blackwell, D., "Discrete dynamic programming," Ann. Math. Statist., vol. 33 (1962), pp. 719-726.

[10] ________, "Discounted dynamic programming," Ibid., vol. 36 (1965), pp. 226-235.

[11] Brown, B. W., "On the iterative method of dynamic programming on a finite space discrete time Markov process," Ann. Math. Statist., vol. 36 (1965), pp. 1279-1285.

[12] Charnes, A. and Cooper, W. W., "Programming with linear fractional functionals," Nav. Res. Log. Q., vol. 9 (1962), pp. 181-186.

[13] Chow, Y. and Cassignol, E., Linear Signal-Flow Graphs and Applications, Wiley, New York, 1962.

[14] Cox, D. R., Renewal Theory, Methuen, London, 1962.

[15] Dantzig, G. B., Linear Programming and Extensions, Princeton University Press, Princeton, 1963.

[16] De Cani, L. S., "A dynamic programming algorithm for embedded Markov chains when the planning horizon is at infinity," Management Sci., vol. 10 (1964), pp. 716-733,

[17] De Ghellinck, G. T., and Eppen, G. D., "Linear programming solutions for separable Markovian decision problems," Ibid., vol. 13 (1967), pp. 371-394.

[18] Denardo, E. V., "Contraction mappings in the theory underlying dynamic programming," SIAM Review, vol. 9 (1967), pp. 165-177.

[19] ________ and Fox, B. L., "Multichain Markov renewal programs," SIAM J. Appl. Math., vol. 16 (1968), pp. 468-487.

[20] Derman, C., "On sequential decisions and Markov chains," Management Sci., vol. 9 (1962), pp. 16-24.

[21] D'Epenoux, F., "A probabilistic production and inventory problems," Management Sci., vol. 10 (1963), pp. 98-108.

[22] Dolazza, E., "System states analysis and flow graph diagrams in reliability," IEEE Trans. Reliability, vol. R-15 (1966), pp. 85-94.

[23] Downton, F., "The reliability of multiplex systems with repair," J. Roy. Statist. Soc., Ser. B, vol. 28 (1966), pp. 459-476.

[24] Dubins, L. E. and Savage, L. J., How to Gamble If You Must, McGraw-Hill, New York, 1965.

[25] Eaton, J. H. and Zadeh, L. A., "Optimal pursuit strategies in discrete-state probabilistic systems," Trans. ASME, Ser. D, J. Basic Engineering, vol. 84 (1962), pp. 23-29.

[26] Epstein, B. and Hosford, H., "Reliability of some two-unit redundant systems," Proc. 6th Nat'l Symp. on Reliability and Quality Control, pp. 466-476, 1960.

[27] Feller, W., An Introduction to Probability Theory and Its Applications, vol. I, 2nd Edition, Wiley, New York, 1957.

[28] ________, An Introduction to Probability Theory and Its Applications, Vol. II, Wiley, New York, 1966.

[29] Flehinger, B. J., "A general model for the reliability analysis of system under various preventive maintenance policies," Ann. Math. Statist., vol. 33 (1962), pp. 137-156.

[30] Flehinger, B. J., "A Markovian model for the analysis of marginal testing on system reliability," Ann. Math. Statist., vol. 33 (1962), pp. 754-766.

[31] Fox, B., "Markov renewal programming by linear fractional programming," SIAM J. Appl. Math., vol. 14 (1966), pp. 1418-1432.

[32] Gaver, Jr., D. P. "Time to failure and availability of paralleled systems with repair," IEEE Trans. Reliability, vol. R-12 (1963), pp. 30-38.

[33] ________, "Failure time for a redundant repairable system of two dissimilar units," Ibid., vol. R-13 (1964), pp. 14-22.

[34] Gnedenko, B. V., "Some theorems on standbys," Proc. Fifth Berkeley Symp. on Math. Statist. and Prob., edited by Le Cam and J. Neyman, vol. III, pp. 285-290, University of California Press, Berkeley, 1967.

[35] ________, Belyaev, Yu. K., and Solovyev, A. D., Mathematical Methods of Reliability Theory, English Translation edited by R. E. Barlow, Academic Press, New York, 1969.

[36] Halperin, M., "Some waiting time distributions for redundant systems with repair," Technometrics vol. 6 (1964), pp. 27-40.

[37] Harris, R., "Reliability applications of a bivariate exponential distributions," Operations Research, vol. 16 (1968), pp. 18-27.

[38] Hoffman, A. J. and Karp, R. M., "On nonterminating stochastic games," Management Sci., vol. 12 (1966), pp. 359-370.

[39] Hosford, J. E., "Measures of dependability," Operations Research, vol. 8 (1960), pp. 53-64.

[40] Howard, R. A., Dynamic Programming and Markov Processes, M. I. T. Press, Cambridge, 1960.

[41] ________, "Research in semi-Markovian decision structures," J. Opns. Res. Soc. Japan, vol. 6 (1964), pp. 163-199.

[42] Huggins, W. H., "Flow graph representation of systems," in Operations Research and Systems Engineering, edited by Fragle, Huggins, and Roy, Chap. 21, The Johns Hopkins University Press, Baltimore, 1962.

[43] ________, "System dynamics," Ibid., Chap. 22, 1962.

[44] Jewell, W. S., "Markov-renewal programming. I. Formulation, finite return model," Operations Research, vol. 11 (1963), pp. 938-948.

[45] ________, "Markov-renewal programming. II. Infinite return model, example," Ibid., vol. 11 (1963), pp. 949-971.

[46] Karlin, S,. "The structure of dynamic programming models," Nav. Res. Log. Q., vol. 2 (1955), pp. 284-294.

[47] Kemeny J. G. and Snell, J. L. Finite Markov Chains, Van Nostrand, Princeton, 1960.

[48] Klein, M., "Inspection-maintenance-replacement schedules under Markovian deterioration," Management Sci., vol. 6 (1960), pp. 259-267.

[49] Liebowitz, B. H., "Reliability considerations for a two element redundant system with generalized repair times," Operations Research, Vol. 14 (1966), pp. 233-241.

[50] Manne, A. S. "Linear programming and sequential decisions," Management Sci., vol. 6 (1960), pp. 259-267.

[51] Mason, S. J., "Feedback theory: Some properties of signal flow-graphs," Proc. IRE, vol. 41 (1953), pp. 1144-1156.

[52] ________, "Feedback theory: Further properties of signal flow-graphs," Ibid., vol. 44 (1956), pp. 920-926.

[53] Pyke, R., "Markov renewal processes: Definitions and preliminary properties," Ann. Math. Statist., vol. 32 (1961), pp. 1231-1242.

[54] ________, "Markov renewal processes with finitely many states," Ibid., vol. 32 (1961), pp. 1243-1259.

[55] Shapley, L. S., "Stochastic games," Proc. Nat. Acad. Sci., vol. 39 (1953), pp. 1095-1100.

[56] Smith, W. L., "Regenerative stochastic processes," Proc. Roy. Soc. London, Ser. A, vol. 232 (1955), pp. 6-31.

[57] ________, "Renewal theory and its ramifications," J. Roy. Statist. Soc., Ser. B, vol. 20 (1958), pp. 243-302.

[58] Srinivasan, V. S., "The effect of standby redundancy in system's failure with repair maintenance," Operations Research, vol. 14 (1966), pp. 1024-1036.

[59] ________, "First emptiness in the spare parts problem for repairable components," Ibid., vol. 16 (1968), pp. 407-415.

[60] ________, "A standby redundant model with non-instantaneous switchover," IEEE Trans. Reliability, vol. R-17 (1968), pp. 175-178.

[61] Strauch, R. E., "Negative dynamic programming," Ann. Math. Statist., vol. 37 (1966), pp. 871-890.

[62] Tin Htun, L., "Reliability prediction techneques for complex systems," IEEE Trans. Reliability, vol. R-15 (1966), pp. 58-69.

[63] Veinott, Jr., A. F., "On the finding optimal policies in discrete dynamic programming with no discounting," Ann. Math. Statist., vol. 37 (1966), pp. 1284-1294.

[64] Wagner, H. M., "On the optimality of pure strategies," Management Sci., vol. 6 (1960), pp. 268-269.

[65] Wolfe, P. and Dantzig, G. B., "Linear programming in a Markov chain," Operations Research, vol. 10 pp. 702-710.